Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US05/008977

International filing date: 18 March 2005 (18.03.2005)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/554,318
Filing date: 18 March 2004 (18.03.2004)

Date of receipt at the International Bureau: 12 May 2005 (12.05.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

1315307

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

April 28, 2005

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/554,318*
FILING DATE: *March 18, 2004*
RELATED PCT APPLICATION NUMBER: *PCT/US05/08977*

Certified by

Jon W. Dudas

Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

20427 U.S. PTO
031804

PTO/SB/16 (08-03)
Approved for use through 07/31/2006. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

Express Mail Label No. EV 414834544 US

INVENTOR(S)		
Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Larry R.	Rohrschneider	Seattle, WA

Additional inventors are being named on the _____ separately numbered sheets attached hereto

TITLE OF THE INVENTION (500 characters max)

Methods and Compositions Involving S-Ship Promoter

22858 U.S. PTO
60/554318
031804

Direct all correspondence to: **CORRESPONDENCE ADDRESS**

☑ Customer Number: 32425

OR

☑ Firm or Individual Name	Gina N. Shishima				
Address	Fulbright & Jaworski L.L.P.				
Address	600 Congress Avenue, Suite 2400				
City	Austin	State	TX	Zip	78701
Country	USA	Telephone	512-536-3081	Fax	512-536-4598

ENCLOSED APPLICATION PARTS *(check all that apply)*

☑ Specification *Number of Pages* 82 + 33-pg. sequence
☐ CD(s), Number _____
☑ Drawing(s) *Number of Sheets* 5
☑ Other (specify) Return postcard
☐ Application Date Sheet. See 37 CFR 1.76

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT

☑ Applicant claims small entity status. See 37 CFR 1.27.
☐ A check or money order is enclosed to cover the filing fees.
☑ The Director is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: 50-1212/FHCC:016USP1
☐ Payment by credit card. Form PTO-2038 is attached.

FILING FEE Amount ($): 80.00

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☑ No.
☐ Yes, the name of the U.S. Government agency and the Government contract number are: _____

Respectfully submitted,

SIGNATURE Michael R. Krawzenek (51,898) Michael R. Krawzenek for:

Date March 18, 2004

REGISTRATION NO. 45,104 *(if appropriate)*

Docket Number: FHCC:016USP1

TYPED or PRINTED NAME Gina N. Shishima

TELEPHONE 512-536-3081

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

This collection of information is required by 37 CFR 1.51. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 8 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: **Mail Stop Provisional Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

PATENT

FHCC:016USP1

PROVISIONAL

APPLICATION FOR UNITED STATES LETTERS PATENT

for

METHODS AND COMPOSITIONS INVOLVING S-SHIP PROMOTER REGIONS

by

Larry R. Rohrschneider

CERTIFICATE OF EXPRESS MAIL	
Express Mail No.:	EV 414834544 US
Date of Deposit:	March 18, 2004

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates generally to the fields of molecular and developmental biology. More particularly, it concerns methods and compositions involving *s-SHIP* promoter regions, which can be used to promote transcription in particular cell types and at particular times during development.

2. Description of Related Art

Stem cells have been the focus of tremendous interest in recent years because of the progress made in developmental and molecular biology and the promise of therapeutic applications in a wide variety of contexts, from heart disease to diabetes, and cancer to Parkinson's disease (see generally Abbott *et al.*, 2003; Daley, 2003; Hirai, 2002; Kondo *et al.*, 2003; Nakano, 2003). Toward fulfilling this promise, many researchers have engaged in extensive studies to characterize factors and pathways in stem cell development and to evaluate candidate therapeutic and diagnostic agents. Such agents include proteins that are gene products, sometimes heterologous, in the stem cells. The ability to express a transgene in stem cells is critical for providing data toward these endeavors. The study of genes normally expressed in stem cells has yielded not only information regarding the developmental, cellular, and molecular biology of these cells, but also useful tools for further studies.

Pathways involved in stem cell function include the protein phosphatidylinositol 3-kinase (PI3K), which becomes activated through cell surface receptors. PI3K is involved in the generation of phosphatidylinositol 3, 4, 5-triphosphate, which activates signaling pathways leading to cell proliferation. The SH2-containing inositol 5'-phosphatase (SHIP1) removes the phosphate group from the D5 position of phosphatidylinositols, which is considered an significant feedback mechanism on cell activation for hematopoietic cells (Lioubin *et al.*, 1996; Rohrschneider *et al.*, 2000).

A form of SHIP1 lacking the SH2 domain has been identified and referred to as stem or short SHIP (s-SHIP) (Tu *et al.*, 2001). Tu *et al.* found the protein contains amino acids encoded by exons 6-27 of *SHIP1* and that it is expressed in embryonic and

hematopoietic stem cells. It was initially unclear how s-SHIP protein was produced from the *ship1* gene. Kavanaugh *et al.* (1996) suggested that SIP-110 was a spliced product of SHIP1; however, Tu *et al.* (2001) proposed, based on its cDNA sequence, that it was transcribed from a promoter within the *SHIP1* gene. This was inferred from the fact that the first 44 nucleotides of the s-SHIP cDNA were at the 5' end immediately before exon 6 of *SHIP1*. These 44 nucleotides were not contained in the SHIP1 cDNA, but were identical to the 44 nucleotides of genomic *ship1* intron 5, immediately adjacent to exon 6. However, no functional evidence for an *s-ship* promoter was provided. Therefore, while a promoter with the tissue-specific expression of *s-ship* could be valuable from both a research and therapeutic/diagnostic perspective, further investigation of the *s-ship* gene was required to identify the promoter and characterize any tissue-specific activity.

SUMMARY OF THE INVENTION

The present invention concerns methods and compositions involving a functional *s-ship* promoter. The invention includes nucleic acid molecules, host cells, and transgenic organisms having an *s-ship* promoter, as well as methods of using the promoter for transcription, expression studies, stem cell analyses, and therapeutic applications.

The present invention concerns an *s-ship* promoter and its functional derivatives. The term "promoter" is used according to its ordinary and plain meaning to a person skilled in the art of eukaryotic transcriptional regulation. The terms "*s-ship* promoter" or "*s-ship1* promoter" refer to the nucleic acid sequence from the *s-ship* gene that is capable of promoting transcription of a nucleic acid sequence that is connected to it (downstream). Transcription can be assayed according to any number of ways known to those of skill in the art, including, but not limited to, an expression assay using a screenable or selectable marker; ribonuclease protection assay (RNAP), RT-PCR, and *in vitro* transcription reactions, all of which are well known to those of skill in the art and can be implemented using commercially available reagents and protocols (see generally, Sambrook *et al.*, 1989; Ausubel, 1992 and 1994, all of which are incorporated by reference).

Compositions of the invention include isolated polynucleotides comprising an *s-ship* promoter capable of promoting transcription. SEQ ID NO:1 is a 102 kb genomic mouse *ship1* sequence. In certain embodiments, the *s-ship1* promoter comprises, or has at least or at most 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 12, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200, 3300, 3400, 3500, 3600, 3700, 3800, 3900, 4000, 4100, 4200, 4300, 4400, 4500, 4600, 4700, 4800, 4900, 5000, 6000, 7000, 8000, 9000, 10000, 11000, 12000, 13000, 14000, 15000, 16000, 17000, 18000, 19000, 20000, 21000, 22000, 23000, 24000, 25000, 26000, 27000, 28000, 29000, 30000, or more contiguous nucleotides of SEQ ID NO:1, or any range derivable therein. In specific embodiments, *s-ship1* promoter includes one or more of the following regions of SEQ ID NO:1: from nucleotide (nt) 49485 to 60914 (11.5 kb-GFP construct); 49485 to 57072, which is 7588 nt (7.6 kb construct); from nt 49485 to 55810, which is 6326 nt (6.3 kb construct); from 49485 to 54755, but lacking 57050 to 57883 (6.2kb-GFP construct); from nt 51389 to 55810, which is 4421 nt (4.4 kb construct); from nt 52199 to 56423, which is 4224 nt (4.2 kb construct); from nt 53820 to 55810, which is 1990 nt (1.9 kb construct); from nt 54755 to 55810, which is 1055 nt (0.96 kb construct); or from nt 55668 to 55810, which is 142 nt (44 nt construct). It is further contemplated that the lengths of contiguous nucleotides discussed above can be applied with respect to these identified regions of SEQ ID NO:1, as well as any other sequence disclosed herein.

It is contemplated that functional derivatives of the *s-ship* promoter also contemplated by the invention. Functional derivatives of an *s-ship* promoter may be at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identical to the polynucleotides discussed herein. Such derivatives may also be characterized by any of the lengths of contiguous nucleotides discussed above.

In specific embodiments, the *s-ship* promoter is capable of promoting tissue-specific transcription. Transcription may be accomplished, in some embodiments of the invention, in skin, a hair follicle, cornea, embryo, gonads, mammary gland, pancreas, and/or smooth muscle. It is also contemplated that transcription may be achieved in cells qualified as or with characteristics of stem cells, which may or may not be derived from skin, a hair follicle, cornea, embryo, gonads, mammary gland, pancreas, and/or smooth muscle. In some embodiments, transcription is achieved in a hematopoietic cell in a tissue-specific manner, including in hematopoietic stem or progenitor cells, but also in more mature or differentiated cells.

In some embodiments of the invention, an *s-ship* promoter is operably attached to a heterologous nucleic acid. The term "heterologous" is used according to its plain and ordinary meaning to a person skilled in the art of molecular biology. It is a relative term and in the context of an *s-ship* promoter, it refers to a nucleic acid sequence that is not normally found in nature (with respect to sequence and position) with the *s-ship* promoter. In other words, it refers to any nucleic acid that is not the entire genomic sequence of the *s-ship* gene. In some embodiments, the *s-ship* promoter is connected to a nucleic acid sequence encoding part of the *s-ship* gene product or all or part of an *s-ship1* cDNA sequence. Alternatively, the *s-ship* promoter may be placed at a location different than is found in nature.

Because recombinant cells and transgenic animals, including knockout versions thereof, are part of the invention, the present invention further encompasses nucleic acids containing an *s-ship* gene or a portion thereof and a marker sequence, wherein the *s-ship* gene is disrupted by the marker sequence. In some embodiments, the nucleic acid is under the control of a promoter, which is an *s-ship* promoter in further embodiments. The promoter may also be constitutive, inducible, or conditional. Promoters discussed herein may be tissue-specific (spatially restricted), developmental-specific (providing transcription at specific developmental stages or times), and/or temporally restricted.

The present invention further concerns expression cassettes, vectors, and host cells that contain or include polynucleotides having an *s-ship* promoter that has been isolated away from its chromosomal context. The polynucleotides and embodiments

discussed above may be implemented with respect to expression cassettes, vectors, and host cells.

It is contemplated that the *s-ship* promoter may control the transcription of a nucleic acid sequence encoding a marker. In some embodiments, the marker is colorimetric, enzymatic, or fluorescent. Examples include, but are not limited to, β-galactosidase, chloramphenicol acetylase, luciferase, and green fluorescent protein. In further embodiments, a heterologous nucleic acid segment encodes a therapeutic or diagnostic gene product. The therapeutic or diagnostic gene product may be a protein or RNA molecule, such as an siRNA or miRNA molecule. In some embodiments, the therapeutic gene product is selected from the group consisting of a tumor suppressor, a cytokine, a cytokine receptor, a differentiation-inducer, growth factor, and a growth factor receptor. Examples of such proteins are well known to those of skill in the art, and include, but are not limited to, interleukins (IL-2, -6, -8, -9, -10, -11, -12, -13, -14, -15, -16, -17, -18, -19, -20, -21, -22, -23, -24, etc.), interferons, receptor tyrosine kinases and their ligands (kit/steel, CSFR/CSF, GM-CSFR/GM-CSF, PDGFR/PDGF, flk-1/VEGF, Lif, EGF, FGF, etc.), transforming growth factors α and β, Epo, IGF, tumor necrosis factor α and β. A number of examples can be seen on the world wide web at indstate.edu/thcme/mwking/growth-factors.html, which is specifically incorporated by reference.

In some embodiments of the invention, a vector is a plasmid, YAC, BAC, or virus. Viruses include adenovirus, adeno-associated virus, retrovirus, flavivirus, and vaccinia virus.

Compositions of the invention may be prepared in a pharmaceutically or pharmacologically acceptable formulation. Such formulations are well known to those of skill in the art for use in *in vivo* contexts.

Other aspects of the invention include host cell having an *s-ship* promoter operably attached to a heterologous nucleic acid segment. In some embodiments, the host cell is eukaryotic, though it may be prokaryotic. In specific embodiments, the host cell is from a mammal, insect, bacteria, or yeast. Cells from monkeys, mice, rats, rabbits, hamsters, ferrets, and humans are specifically contemplated for use with nucleic acids of the invention. In some cases, the host cell is an embryonic cell, which may specifically be

a blastocyst cell. In other cases, the host cell is a stem or progenitor cell. In some cases, the cell is a hematopoietic cell, meaning any cell in that lineage. It is contemplated that the cell may be *in vitro* or *in vivo*.

Cells that can be used according to methods and compositions of the invention include, but are not limited to, CD34+ cells (cells expressing CD34 on their surface), undifferentiated cells, stem cells, progenitor cells, cord blood cells, placental cells, neonatal or fetal cells, immature cells, pluripotent cells, and totipotent cells. The term "stem cell" is used according to its ordinary meaning, for example, as described by the National Institutes of Health (on the World Wide Web at stemcells.nih.gov). Stem cells 1) are "capable of dividing and renewing themselves for long periods"; 2) are unspecialized; and, 3) can give rise to specialized cell types.

The invention specifically contemplates the use of embryonic stem cells, adult stem cells, or neonatal and fetal stem cells. An adult stem cell typically refers to a stem cell from a particular organ or tissue that is capable of differentiating into one or more cells of that organ or tissue. Umbilical cord blood contains stem cells that are similar to embryonic stem cells in that they are believed to be capable of being differentiated into a number of different cell types, as opposed to cell types of one particular organ or tissue. Umbilical cord blood refers to blood that remains in the umbilical cord and placenta following birth and after the cord is cut. "Placental blood" is understood to be synonymous with cord blood; similarly, cord blood stem cell is considered synonymous with placental or placental blood stem cell. The use of stem cells from umbilical cord blood is specifically contemplated in certain embodiments of the invention. In some but not all cases, the use of other stem cells is specifically not considered part of the invention, particularly the use of pancreatic/endocrine progenitor or stem cells is not considered for use with some embodiments. Furthermore, cells of the invention may be characterized by cell surface antigens. Cell surface antigens and their correlation with cell type and cell development are known to those of ordinary skill in the art.

It will be understood that cultures or samples containing cells discussed above are also contemplated for use according to methods and compositions of the invention.

Further embodiments of the invention include cells for use in the generation of transgenic organisms (knock-in and knock-out). Accordingly, there are recombinant host

cells in which one or both *s-ship* genes is disrupted by marker sequence or in which all or part of an *s-ship* gene is flanked by an excisable sequence, such as a loxP sequence. The marker sequence serves the purpose of showing when the transgenic sequence is present or absent in the cell.

The present invention further concerns transgenic animals comprising an *s-ship* promoter operably attached to a heterologous nucleic acid segment. Mammals are specifically contemplated, particularly mice. In some cases, the invention involves a mammal having cells comprising an *s-ship* transgenic sequence. The *s-ship* sequence may be knocked in or out in a restricted or controlled manner. For example, whether it is knocked in or out may be controlled in a tissue-specific, inducible, conditional, developmental or temporal manner. Consequently, animals may have heterologous genes under the control of a promoter or system that operates in that way. The Cre-Lox system is one example. The transgene of interest itself may not be under the control of a limited promoter, but a secondary gene whose product initiates the knock-in or knock-out process may be under such a promoter. In one embodiments, animals of the invention may have an *s-ship* transgenic sequence that includes an *s-ship* coding sequence flanked by loxP sequences. They may also have a heterologous nucleic acid sequence encoding a Cre recombinase. In some cases, the nucleic acid sequence encoding the Cre recombinase is under the control of an inducible or conditional promoter. Transgenic animals of the invention are not limited by the Cre-Lox system, which serves as an example of how expression may be controlled.

A number of methods are included as part of the present invention. In some embodiments, there are methods for expressing a recombinant nucleic acid in a cell comprising:a) transfecting the cell with an expression cassette comprising an *s-ship* promoter operably attached to the recombinant nucleic acid, wherein the nucleic acid is transcribed. The cell may be any of the host cells discussed above.

Other embodiments of the invention concern methods of screening for a candidate substance that regulates activity of the *s-ship* promoter comprising a step selected from the group consisting of: (a) contacting a nucleic acid comprising an *s-ship* promoter with an *s-ship* promoter binding protein and the candidate substance under conditions that allow binding between the protein and the promoter and determining whether the

candidate compound modulates the binding between the protein and the promoter; and (b) contacting the candidate substance with a cell comprising the *s-ship* promoter operably attached to a reporter gene coding for an expression product and assaying for expression of the reporter gene expression product. One or both steps may be employed. Ways of determining whether the candidate compound modulates binding between a protein and the promoter are well known to those of skill in the art. The compound may inhibit, reduce, decrease, eliminate, increase, promote, tighten the binding between the protein and the promoter. Assays for such an interaction include, but are not limited to, electrophoretic mobility shift assays (EMSA), DNA footprinting, functional transcription assays—as described above—Southwestern assays, and PCR-based assays.

The present invention also includes methods for identifying stem cells in a population of cells comprising: (a) administering to cells in the population a nucleic acid comprising an *s-ship* promoter operably attached to a reporter or marker gene. The reporter or marker gene is then used to identify positively-expressing cells, which would indicate the cell is a stem cell. The cell may be in an organ and/or in an animal. In some embodiments, methods include sorting cells based on expression of the reporter or marker gene. In addition to the assays discussed above, FACS analysis may be employed, in addition to other cell sorting techniques. Methods include differentiation of the cells.

Aspects of the invention also concern methods for screening for a modulator of cell function comprising: a) transfecting a stem or hematopoietic cell with an expression cassette comprising an *s-ship* promoter operably attached to a nucleic acid encoding a candidate modulator; and, b) assaying the cell for a cell function, wherein a difference in cell function in the cell as compared to a cell in the absence of the candidate modulator is indicative of a modulator. The term "modulator" refers to a substance that affects cell function. It may affect cell function by acting on or through a pathway. The modulator may inhibit, reduce, eliminate, decrease, increase, promote, induce, or enhance a particular cell function or result of a pathway in the cell. It is contemplated that this method may be employed to identify a modulator as a candidate therapeutic agent for the treatment of a blood-related disease or condition.

Therapeutic methods are also provided by the present invention. Methods are not necessarily limited to a particular disease or condition. It is contemplated that any method

in which expression in stem cells or cells in which the *s-ship* promoter can function are contemplated for use in therapeutic methods of the invention. For example, the method may be applied to pancreatic disorders and diseases.

Thus, in some embodiments, there is a method of treating a patient with a blood-related disease or condition comprising: a) transfecting a cell with an expression cassette comprising an *s-ship* promoter region operably attached to a therapeutic nucleic acid; and, b) administering the cell to the patient. Blood-related disease or condition include blood-related cancers—such as leukemia, lymphoma, or myeloma—and anemia. In some cases, the blood-related condition can be treated using stem cell replacement therapy.

Cells for therapeutic use may, in addition to the cells discussed above, be bone marrow cells, or be autologous or allogeneic.

It is contemplated that any embodiment discussed with respect to any method or composition described herein can be implemented with respect to any other method or composition described herein.

The use of the term "or" in the claims is used to mean "and/or" unless explicitly indicated to refer to alternatives only or the alternatives are mutually exclusive, although the disclosure supports a definition that refers to only alternatives and "and/or."

Throughout this application, the term "about" is used to indicate that a value includes the standard deviation of error for the device or method being employed to determine the value.

Following long-standing patent law convention, the words "a" and "an," when used in conjunction with the word "comprising" in the claims or specification, denotes one or more, unless specifically noted.

Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples, while indicating specific embodiments of the invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

FIG 1. *ship1* genomic segments cloned into a promoter-less expression vector for testing cell-specific promoter activity. The upper line represents the general *ship1* genomic region containing potential activity for cell-specific s-SHIP expression. Intron 5 contains the likely promoter activity and transcription is proposed to begin before exon 6. The 44-intronic nucleotides, contained in the s-SHIP cDNA, are shown as red. A 7.6 kb genomic fragment (second line down), as well as the indicated sub-fragments, were cloned into a promoter-less plasmid for GFP expression. The design and construction of the plasmid is detailed in Materials and Methods.

FIG. 2. Flow cytometry analysis of cell type-specific promoter activity in D3 ES cells vs. NIH3T3 cells. Each construct shown was cloned into a promoter-less GFP plasmid, which was linearized and electroporated into D3 ES cells, or transfected into NIH3T3 cells. G418 resistant colonies were then examined by flow cytometry for GFP expression. Two different "empty vector" negative controls were utilized depending on whether the insert contained a splice acceptor or both splice acceptor and donor. Both these plasmids without genomic insert were negative for GFP expression in both cell types, but only a single negative-control vector is shown. Two positive-control plasmids were utilized in each experiment. These controls expressed GFP from an IRES, and one expressed a protein insert, both were positive in each cell type.

FIG. 3A-B. Structure of the 11.5-kb and 6.2-kb transgenic promoter-GFP constructs for in vivo analysis. **FIG. 3A**. Two promoter transgenic constructs were prepared. The first construct is called the 11.5kb-GFP transgene, and contains the entire genomic *ship1* segment from the Sac I site near the 5' end of intron 5 through the putative translation start site at an ATG preceded by a suitable Kozac sequence within exon 7. The translational start ATG for the enhanced GFP is fused, in frame, to the likely

ATG translational start for s-SHIP. A second transgenic construct, called the 6.2kb-GFP transgene, is identical to the 11.5kb-GFP construct, except it contains only 0.96 kb upstream of exon 6, and lacks 833 nt within intron 6. **FIG. 3B**. Transgenic copy numbers were estimated by semi-quantitative PCR analysis relative to the endogenous diploid *gab2* gene.

FIG. 4. Computer analysis of 600 nucleotides of the intron-5 transgene promoter region. A. The region immediately upstream of exon 6 is shown with potential transcription factor binding motifs determined by analysis using the MatInspector program. Only the factors with matrix and core similarity greater than 0.9 are shown. Those factor motifs within the strand shown are over-lined, while those factors potentially interacting with the complementary strand are shown underlined. The SSR or stem-SHIP region identified by Tu *et al.*, 2001 is in bold, and an initiator sequence for transcription is situated at the beginning of the SSR. Exon 6 (not shown) begins at the 3' end of the SSR.

FIG. 5. The two primary proteins, s-SHIP and SHIP1, are produced from the *ship1* gene. The domain structure of the two proteins is shown above the *ship1* genomic intron/exon organization. Transcription for the full-length 145 kDa SHIP1 protein initiates in promoter 1 (Pro1), utilizes all 27 exons, and translation begins in the exon 1 encoded region. The stop codon is the first three nucleotides of exon 27. Transcription for the s-SHIP protein begins within intron 5 (Pro 2), and downstream is identical to the SHIP1 product. Translation, however, presumably begins in the first ATG of exon 7. Both transcripts and protein sequences are therefore identical from the ATG in exon 7 through the stop codon in exon 27. The dashed lines indicate translation start and stop points for each protein within the genomic exons.

DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

The present invention is based on the isolation and characterization of the *s-ship1* promoter, which can be used to promote transcription. Methods and compositions involving the *s-ship1* promoter are provided herein. In some embodiments, they take advantage of the tissue specificity of *s-ship1* expression. *s-ship1* encodes a protein whose

expression has been observed in limited cell populations, and thus, the tissue-specificity of its promoter can be exploited in a number of different ways.

I. SHIP1 and s-SHIP Background

The *s-ship1* promoter was studied because of the function and expression patterns for the *s-ship1* (also referred to as *s-ship*) and *ship1* gene products. The murine SHIP1 protein is encoded in 27 exons of the Inpp5d (inositol polyphosphate-5-phosphatase D) locus, spanning approximately 102 kbps on chromosome 1 at position 57.0 cM of the genetic map, or cytoband C5 of the cytogenetic map (reviewed in Rohrschneider *et al.*, 2000; Wolf *et al.*, 2001; NCBI databases). The full-length protein is 145 kDa, but splicing, involving exons 25 and 26, can produce 4 additional proteins ranging in size from 109-135 kDa (Lucas and Rohrschneider, 1999; Wolf *et al.*, 2000). These splicing reactions affect the 350-amino acid C-terminal tail region and its numerous protein interaction motifs, such as those binding PTB, SH2, and SH3 domains.

The prominent structural features of the SHIP1 protein dictate its major functional interactions. The SHIP1 SH2 domain has general specificity for tyrosine-phosphorylated Yxx(L/I/V) amino acid motifs, and its inositol 5'phosphatase domain removes phosphate from the 5' position of inositol(3,4,5)P_3, phosphatidylinositol(3,4,5)P_3 or Inositol(1,3,4,5)P_4 [see Sly *et al.*, (2003) for review]. The tyrosine-phosphorylated C-terminal tail interacts directly with the PTB domain of Shc and Dok proteins (Lioubin *et al.*, 1996; Sattler *et al.*, 2001; Tamir *et al.*, 2000), and a potential interaction motif for the SH2 domain of the p85 component of the p85/PI3K is present in the full-length SHIP1 (Gupta *et al.*, 1999; Lucas and Rohrschneider 1999), but eliminated by the splicing events producing the α and β isoforms (Rohrschneider *et al.*, 2000). Polyproline-rich interaction motifs for the SH3 domains of Grb2 also are present in the C-tail region (Kavanaugh *et al.*, 1996). The SHIP1 proteins (*e.g.*, the 145 kDa protein and isoforms thereof) are expressed in hematopoietic cells and testes, with lower expression observed in a few other adult tissues (Q. Liu *et al.*, 1998, reviewed in Rohrschneider, 2003).

Functionally, both biochemical and genetic experiments indicate SHIP1 is a negative regulator of myeloid cell proliferation, survival, and perhaps chemotaxis (see Sly *et al.*, 2003; Rohrschneider, 2003). Also, SHIP1 negatively regulates degranulation,

inflammatory cytokine release, and adhesion for mast cells, and SHIP1 is a component of negative signaling (anergy) in B cell proliferation. The molecular mechanisms for most of these effects require the attachment of the SHIP1 SH2 domain to the cytoplasmic portions of transmembrane receptors containing appropriate tyrosine-phosphorylated interaction motifs. There, the SHIP1 inositol-5'phosphatase domain converts the plasma membrane PI3K-produced substrate, phosphatidylinositol(3,4,5)P_3 to phosphatidylinositol(3,4)P_2 effectively terminating proliferation signals. Therefore, the SH2 domain of SHIP1 plays a critical role in initiating many of these negative biological effects.

An additional smaller protein from the *ship1* locus is described as an SH2-less 104 kDa protein (Tu *et al.*, 2001). This product is called s-SHIP, with the prefix signifying its only known expression within two stem cell types (*i.e.*, ES cells and lineage-depleted Sca1-positive cells of the bone marrow). This protein was first described by Kavanaugh *et al.* (1996) and called SIP-110 in the human; but details of its existence were unclear until Tu *et al.* (2001) defined the cDNA and demonstrated endogenous expression in the two cell types described above. Thus, structurally, s-SHIP differs from SHIP1 only by the lack of the N-terminal SH2 domain; but biochemically, s-SHIP also lacks tyrosine phosphorylation and association with Shc (Kavanaugh *et al.*, 1996; Tu *et al.*, 2001). Nevertheless, s-SHIP constitutively interacts with Grb2. The lack of an SH2 domain in s-SHIP indicates its interaction mechanism with target proteins probably differs from that of SHIP1; however, the biological functions of s-SHIP are not known.

II. Nucleic Acids

A. Polynucleotides

The *s-ship1* promoter was identified as a strong promoter for s-SHIP by analyses of the genomic *ship1* intron-5 region in driving GFP expression both *in vitro* and *in vivo*. This promoter exhibited cell-type specific expression in ES cells, and mice transgenic for the promoter (the 11.5kb-GFP transgene) showed tissue-specific GFP expression within the inner cell mass of the blastocyst. Transgenic mice produced with a shorter promoter construct (the 6.2kb-GFP transgene) expressed GFP throughout the blastocyst, suggesting the absence of negative regulatory regions in the shorted transgene. RT-PCR analysis

demonstrated s-SHIP expression within the blastocyst. These results indicate that the 11.5-kb promoter region of the transgene contains the information for tissue-specific expression of *s-SHIP*, as well as tissue-specific shut-off of this protein. It is specifically contemplated that this promoter and the transgenic mice will be useful for future examination of GFP-expression in potential stem/progenitor cells of the embryo and the adult mouse.

The present invention concerns polynucleotides, isolatable from cells, that are free from total genomic DNA and that contain an *s-ship* promoter. It is contemplated that the *s-ship1* promoter is capable of directing transcription of nucleic acid sequence. Transcription may be directed in a tissue-specific or developmental manner. The nucleic acid sequence may encode a peptide or polypeptide, or it may also encode an RNA molecule that is not translated into a protein.

A "promoter" is a control sequence that is a region of a nucleic acid sequence at which initiation and rate of transcription are controlled. It may contain genetic elements at which regulatory proteins and molecules may bind, such as RNA polymerase and transcription factors. The phrases "operatively positioned," "operatively linked," "under control," "operatively attached," and "under transcriptional control" mean that a promoter is in a correct functional location and/or orientation in relation to a nucleic acid sequence to control transcriptional initiation and/or expression of that sequence. Typically, the promoter is located 5' or upstream from the strand of sequence to be transcribed. A promoter may or may not be used in conjunction with an "enhancer," which refers to a cis-acting regulatory nucleic acid sequence involved in the transcriptional activation of a nucleic acid sequence.

As used herein, the term "DNA segment" or "nucleic acid segment" refers to a DNA or nucleic acid molecule that has been isolated free of total genomic DNA of a particular species. Therefore, a DNA segment encoding a polypeptide refers to a DNA segment that contains wild-type, polymorphic, or mutant polypeptide-coding sequences yet is isolated away from, or purified free from, total mammalian or human genomic DNA. Included within the term "DNA segment" are a polypeptide or polypeptides, DNA segments smaller than a polypeptide, and recombinant vectors, including, for example, plasmids, cosmids, phage, viruses, and the like.

As used in this application, the term "s-ship polynucleotide" refers to an s-ship-encoding nucleic acid molecule. The term "cDNA" is intended to refer to DNA prepared using messenger RNA (mRNA) as template.

It also is contemplated that a particular polypeptide from a given species may be represented by natural variants that have slightly different nucleic acid sequences but, nonetheless, encode the same protein.

Similarly, a polynucleotide comprising an isolated or purified wild-type, polymorphic, or mutant polypeptide gene refers to a DNA segment including wild-type, polymorphic, or mutant polypeptide coding sequences isolated substantially away from other naturally occurring genes or protein encoding sequences. In this respect, the term "gene" is used for simplicity to refer to a functional protein, polypeptide, or peptide-encoding unit. As will be understood by those in the art, this functional term includes genomic sequences, cDNA sequences, and smaller engineered gene segments that express, or may be adapted to express, proteins, polypeptides, domains, peptides, fusion proteins, and mutants. A nucleic acid encoding all or part of a native or modified polypeptide may contain a contiguous nucleic acid sequence encoding all or a portion of such a polypeptide of the following lengths: about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 441, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000, 1010, 1020, 1030, 1040, 1050, 1060, 1070, 1080, 1090, 1095, 1100, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 9000, 10000, or more nucleotides, nucleosides, or base pairs.

In particular embodiments, the invention concerns isolated DNA segments and recombinant vectors incorporating an *s-ship* promoter with a heterologous nucleic acid sequence or a ship or s-ship cDNA segment. Thus, an isolated DNA segment or vector containing a DNA segment may encode, for example, the heterologous nucleic acid sequence. The term "recombinant" may be used in conjunction with a polypeptide, the name of a specific polypeptide, a nucleic acid sequence, or a host cell, and this generally

means that the entity involves or involved a nucleic acid molecule that was manipulated *in vitro* using recombinant DNA technology.

The nucleic acid segments used in the present invention, regardless of the length of the coding sequence itself, may be combined with other nucleic acid sequences, such as enhancers, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol.

It is contemplated that the nucleic acid constructs of the present invention may encode full-length polypeptide from any source or encode a truncated version of the polypeptide, such that the transcript of the coding region represents the truncated version. The truncated transcript may then be translated into a truncated protein. Alternatively, a nucleic acid sequence may encode a full-length polypeptide sequence with additional heterologous coding sequences, for example to allow for purification of the polypeptide, transport, secretion, post-translational modification, or for therapeutic benefits such as targetting or efficacy. As discussed above, a tag or other heterologous polypeptide may be added to the modified polypeptide-encoding sequence, wherein "heterologous" refers to a polypeptide that is not the same as the modified polypeptide.

In a non-limiting example, one or more nucleic acid constructs may be prepared that include a contiguous stretch of nucleotides of sequences disclosed herein, including the *s-ship* promoter.

A nucleic acid construct may be at least 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 15,000, 20,000, 30,000, 50,000, 100,000, 250,000, 500,000, 750,000, to at least 1,000,000 nucleotides in length, as well as constructs of greater size, up to and including chromosomal sizes (including all intermediate lengths and intermediate ranges), given the advent of nucleic acids constructs such as a yeast artificial chromosome are known to those of ordinary skill in the art. It will be readily understood that "intermediate lengths" and

"intermediate ranges," as used herein, means any length or range including or between the quoted values (*i.e.*, all integers including and between such values).

It is specifically contemplated that nucleic acids of the invention may include, be at most, or be at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200, 3300, 3400, 3500, 3600, 3700, 3800, 3900, 4000, 4100, 4200, 4300, 4400, 4500, 4600, 4700, 4800, 4900, 5000, 5100, 5200, 5300, 5400, 5500, 5600, 5700, 5800, 5900, 6000, 6100, 6200, 6300, 6400, 6500, 6600, 6700, 6800, 6900, 7000, 7100, 7200, 7300, 7400, 7500, 7600, 7700, 7800, 7900, 8000, 8100, 8200, 8300, 8400, 8500, 8600, 8700, 8800, 8900, 9000, 9100, 9200, 9300, 9400, 9500, 9600, 9700, 9800, 9900, 10000, 10100, 10200, 10300, 10400, 10500, 10600, 10700, 10800, 10900, 11000, 11100, 11200, 11300, 11400, 11500, 11600, 11700, 11800, 11900, 12000 or more contiguous nucleotides (or any range derivable therein) of nucleic acid disclosed in this application, including, but not limited to SEQ ID NO:1, intron 5 of the mouse *s-ship* gene, an *s-ship* promoter, and any other SEQ ID NOs.

The various probes and primers designed around the nucleotide sequences of the present invention may be of any length. By assigning numeric values to a sequence, for example, the first residue is 1, the second residue is 2, *etc.*, an algorithm defining all primers can be proposed:

n to n + y

where n is an integer from 1 to the last number of the sequence and y is the length of the primer minus one, where n + y does not exceed the last number of the sequence. Thus, for a

10-mer, the probes correspond to bases 1 to 10, 2 to 11, 3 to 12 ... and so on. For a 15-mer, the probes correspond to bases 1 to 15, 2 to 16, 3 to 17 ... and so on. For a 20-mer, the probes correspond to bases 1 to 20, 2 to 21, 3 to 22 ... and so on.

It also will be understood that this invention is not limited to the particular nucleic acid sequences of SEQ ID NO:1. Recombinant vectors and isolated DNA segments may therefore variously include coding regions, coding regions bearing selected alterations or modifications in the basic coding region, or they may encode biologically functional equivalent sequences. For example, mutations can be made to SEQ ID NO:1 that potentially enhance or alter function relative to the native sequence or alternatively, may be silent with regard to function.

The *s-ship* promoter sequence of the invention is exemplified by the nucleic acid sequence given in SEQ ID NO:1. However, in addition to the unmodified *s-ship* promoter sequence of SEQ ID NO:1, the current invention includes derivatives of this sequence and compositions made therefrom. In particular, the present disclosure provides the teaching for one of skill in the art to make and use derivatives of the *s-ship* promoter. For example, the disclosure provides the teaching for one of skill in the art to delimit the functional elements within the *s-ship* promoter and to delete any non-essential elements. Functional elements also could be modified to increase the utility of the sequences of the invention for any particular application. For example, a functional region within the *s-ship* promoter of the invention could be modified to cause or increase tissue-specific expression. Such changes could be made by site-specific mutagenesis techniques, for example, as described below.

One efficient means for preparing such derivatives comprises introducing mutations into the sequences of the invention, for example, the sequence given in SEQ ID NO:1. Such mutants may potentially have enhanced or altered function relative to the native sequence or alternatively, may be silent with regard to function.

Mutagenesis may be carried out at random and the mutagenized sequences screened for function in a trial-by-error procedure. Alternatively, particular sequences that provide the *s-ship* promoter with desirable expression characteristics could be identified and these or similar sequences introduced into other related or non-related sequences via mutation. Similarly, non-essential elements may be deleted without

significantly altering the function of the elements. It further is contemplated that one could mutagenize these sequences in order to enhance their utility in expressing transgenes in a particular cell type, for example, in a particular stem cell.

The means for mutagenizing a DNA segment containing an *s-ship* promoter sequence of the current invention are well-known to those of skill in the art. Mutagenesis may be performed in accordance with any of the techniques known in the art, such as, but not limited to, synthesizing an oligonucleotide having one or more mutations within the sequence of a particular regulatory region. In particular, site-specific mutagenesis is a technique useful in the preparation of promoter mutants, through specific mutagenesis of the underlying DNA. The technique further provides a ready ability to prepare and test sequence variants, for example, incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into the DNA. Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Typically, a primer of about 17 to about 75 nucleotides or more in length is preferred, with about 10 to about 25 or more residues on both sides of the junction of the sequence being altered.

In general, the technique of site-specific mutagenesis is well known in the art, as exemplified by various publications. As will be appreciated, the technique typically employs a phage vector which exists in both a single stranded and double stranded form. Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage are readily commercially available and their use is generally well known to those skilled in the art. Double stranded plasmids also are routinely employed in site directed mutagenesis which eliminates the step of transferring the gene of interest from a plasmid to a phage.

Site-directed mutagenesis in accordance herewith typically is performed by first obtaining a single-stranded vector or melting apart of two strands of a double-stranded vector which includes within its sequence a DNA sequence that includes the *s-ship* promoter. An oligonucleotide primer bearing the desired mutated sequence is prepared,

generally synthetically. This primer is then annealed with the single-stranded vector, and subjected to DNA polymerizing enzymes such as the *E. coli* polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original non-mutated sequence and the second strand bears the desired mutation. This heteroduplex vector is then used to transform or transfect appropriate cells, such as *E. coli* cells, and cells are selected which include recombinant vectors bearing the mutated sequence arrangement. Vector DNA can then be isolated from these cells and used for plant transformation. A genetic selection scheme was devised by Kunkel *et al.* (1987) to enrich for clones incorporating mutagenic oligonucleotides. Alternatively, the use of PCR™ with commercially available thermostable enzymes such as *Taq* polymerase may be used to incorporate a mutagenic oligonucleotide primer into an amplified DNA fragment that can then be cloned into an appropriate cloning or expression vector. The PCR™-mediated mutagenesis procedures of Tomic *et al.* (1990) and Upender *et al.* (1995) provide two examples of such protocols. A PCR™ employing a thermostable ligase in addition to a thermostable polymerase also may be used to incorporate a phosphorylated mutagenic oligonucleotide into an amplified DNA fragment that may then be cloned into an appropriate cloning or expression vector.

The preparation of sequence variants of the selected promoter DNA segments using site-directed mutagenesis is provided as a means of producing potentially useful promoter sequences and is not meant to be limiting as there are other ways in which sequence variants of DNA sequences may be obtained. For example, recombinant vectors encoding the desired promoter sequence may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants.

As used herein, the term "oligonucleotide-directed mutagenesis procedure" refers to template-dependent processes and vector-mediated propagation which result in an increase in the concentration of a specific nucleic acid molecule relative to its initial concentration, or in an increase in the concentration of a detectable signal, such as amplification. As used herein, the term "oligonucleotide directed mutagenesis procedure" also is intended to refer to a process that involves the template-dependent extension of a primer molecule. The term template-dependent process refers to nucleic

acid synthesis of an RNA or a DNA molecule wherein the sequence of the newly synthesized strand of nucleic acid is dictated by the well-known rules of complementary base pairing (see, for example, Watson and Ramstad, 1987). Typically, vector mediated methodologies involve the introduction of the nucleic acid fragment into a DNA or RNA vector, the clonal amplification of the vector, and the recovery of the amplified nucleic acid fragment. Examples of such methodologies are provided by U.S. Patent No. 4,237,224, specifically incorporated herein by reference in its entirety. A number of template dependent processes are available to amplify the target sequences of interest present in a sample, such methods being well known in the art and specifically disclosed herein below.

One efficient, targeted means for preparing mutagenized promoters or enhancers relies upon the identification of putative regulatory elements within the target sequence. This can be initiated by comparison with, for example, promoter sequences known to be expressed in a similar manner. Sequences which are shared among elements with similar functions or expression patterns are likely candidates for the binding of transcription factors and are thus likely elements which confer expression patterns. Confirmation of these putative regulatory elements can be achieved by deletion analysis of each putative regulatory region followed by functional analysis of each deletion construct by assay of a reporter gene which is functionally attached to each construct. As such, once a starting promoter or intron sequence is provided, any of a number of different functional deletion mutants of the starting sequence could be readily prepared.

As indicated above, deletion mutants of the *s-ship* promoter also could be randomly prepared and then assayed. With this strategy, a series of constructs are prepared, each containing a different portion of the clone (a subclone), and these constructs are then screened for activity. A suitable means for screening for activity is to attach a deleted promoter construct to a selectable or screenable marker, and to isolate only those cells expressing the marker protein. In this way, a number of different, deleted promoter constructs are identified which still retain the desired, or even enhanced, activity. The smallest segment which is required for activity is thereby identified through comparison of the selected constructs. This segment may then be used for the construction of vectors for the expression of exogenous protein.

1. Vectors

Promoter sequences of the invention may be comprised in a vector. The term "vector" is used to refer to a carrier nucleic acid molecule into which a nucleic acid sequence can be inserted for introduction into a cell where it can be replicated. A nucleic acid sequence can be "exogenous," which means that it is foreign to the cell into which the vector is being introduced or that the sequence is homologous to a sequence in the cell but in a position within the host cell nucleic acid in which the sequence is ordinarily not found. Vectors include plasmids, cosmids, viruses (bacteriophage, animal viruses, and plant viruses), and artificial chromosomes (*e.g.*, YACs). One of skill in the art would be well equipped to construct a vector through standard recombinant techniques, which are described in Sambrook *et al.*, (1989) and Ausubel *et al.*, 1996, both incorporated herein by reference. In addition to encoding a polypeptide, a vector may encode other polypeptide sequences such as a tag or targetting molecule. Useful vectors encoding such fusion proteins include pIN vectors (Inouye *et al.*, 1985), vectors encoding a stretch of histidines, and pGEX vectors, for use in generating glutathione S-transferase (GST) soluble fusion proteins for later purification and separation or cleavage. A targetting molecule is one that directs the modified polypeptide to a particular organ, tissue, cell, or other location in a subject's body.

The term "expression vector" refers to a vector containing a nucleic acid sequence coding for at least part of a gene product capable of being transcribed. In some cases, RNA molecules are then translated into a protein, polypeptide, or peptide. In other cases, these sequences are not translated, for example, in the production of antisense molecules, siRNA molecules or miRNA molecules. In addition to *s-ship* promoter regions, expression vectors can contain a variety of other "control sequences," which refer to nucleic acid sequences necessary for the transcription and possibly translation of an operably linked coding sequence in a particular host organism. In addition to control sequences that govern transcription and translation, vectors and expression vectors may contain nucleic acid sequences that serve other functions as well and are described *infra*.

In certain embodiments of the invention, the expression vector comprises a virus or engineered vector derived from a viral genome. The ability of certain viruses to enter cells via receptor-mediated endocytosis, to integrate into host cell genome and express

viral genes stably and efficiently have made them attractive candidates for the transfer of foreign genes into mammalian cells (Ridgeway, 1988; Nicolas and Rubenstein, 1988; Baichwal and Sugden, 1986; Temin, 1986). The first viruses used as gene vectors were DNA viruses including the papovaviruses (simian virus 40, bovine papilloma virus, and polyoma) (Ridgeway, 1988; Baichwal and Sugden, 1986) and adenoviruses (Ridgeway, 1988; Baichwal and Sugden, 1986). These have a relatively low capacity for foreign DNA sequences and have a restricted host spectrum. Furthermore, their oncogenic potential and cytopathic effects in permissive cells raise safety concerns. They can accommodate only up to 8 kb of foreign genetic material but can be readily introduced in a variety of cell lines and laboratory animals (Nicolas and Rubenstein, 1988; Temin, 1986).

The retroviruses are a group of single-stranded RNA viruses characterized by an ability to convert their RNA to double-stranded DNA in infected cells; they can also be used as vectors. Other viral vectors may be employed as expression constructs in the present invention. Vectors derived from viruses such as vaccinia virus (Ridgeway, 1988; Baichwal and Sugden, 1986; Coupar *et al.,* 1988) adeno-associated virus (AAV) (Ridgeway, 1988; Baichwal and Sugden, 1986; Hermonat and Muzycska, 1984) and herpesviruses may be employed. They offer several attractive features for various mammalian cells (Friedmann, 1989; Ridgeway, 1988; Baichwal and Sugden, 1986; Coupar *et al.,* 1988; Horwich *et al.,* 1990).

a. Promoters and Enhancers

A promoter may be one naturally associated with a gene or sequence, as may be obtained by isolating the 5′ non-coding sequences located upstream of the coding segment and/or exon. Such a promoter can be referred to as "endogenous." Similarly, an enhancer may be one naturally associated with a nucleic acid sequence, located either downstream or upstream of that sequence. Alternatively, certain advantages will be gained by positioning the coding nucleic acid segment under the control of a recombinant or heterologous promoter, which refers to a promoter that is not normally associated with a nucleic acid sequence in its natural environment. A recombinant or heterologous enhancer refers also to an enhancer not normally associated with a nucleic acid sequence

in its natural environment. Such promoters or enhancers may include promoters or enhancers of other genes, and promoters or enhancers isolated from any other prokaryotic, viral, or eukaryotic cell, and promoters or enhancers not "naturally occurring," *i.e.*, containing different elements of different transcriptional regulatory regions, and/or mutations that alter expression. In addition to producing nucleic acid sequences of promoters and enhancers synthetically, sequences may be produced using recombinant cloning and/or nucleic acid amplification technology, including PCR™, in connection with the compositions disclosed herein (see U.S. Patent 4,683,202, U.S. Patent 5,928,906, each incorporated herein by reference). Furthermore, it is contemplated the control sequences that direct transcription and/or expression of sequences within non-nuclear organelles such as mitochondria, chloroplasts, and the like, can be employed as well.

Naturally, it may be important to employ a promoter and/or enhancer that effectively directs the expression of the DNA segment in the cell type, organelle, and organism chosen for expression. Those of skill in the art of molecular biology generally know the use of promoters, enhancers, and cell type combinations for protein expression, for example, see Sambrook *et al.* (1989), incorporated herein by reference. The promoters employed may be constitutive, tissue-specific, inducible, and/or useful under the appropriate conditions to direct high level expression of the introduced DNA segment, such as is advantageous in the large-scale production of recombinant proteins and/or peptides. The promoter may be heterologous or endogenous.

In addition to the *s-ship* promoter, other elements/promoters may be employed, in the context of the present invention, to regulate the expression of a gene. Table 1 is a list of other promoters and enhancers that may be used in conjunction with the *s-ship* promoter of the invention; this list also identifies references that indicate how promoters can be evaluated. It is not intended to be exhaustive of all the possible elements involved in the promotion of expression but, merely, to be exemplary thereof. Table 2 provides examples of inducible elements, which are regions of a nucleic acid sequence that can be activated in response to a specific stimulus.

TABLE 1 Promoter and/or Enhancer	
Promoter/Enhancer	References
Immunoglobulin Heavy Chain	Banerji *et al.*, 1983; Gilles *et al.*, 1983; Grosschedl *et al.*, 1985; Atchinson *et al.*, 1986, 1987; Imler *et al.*, 1987; Weinberger *et al.*, 1984; Kiledjian *et al.*, 1988; Porton *et al.*; 1990
Immunoglobulin Light Chain	Queen *et al.*, 1983; Picard *et al.*, 1984
T-Cell Receptor	Luria *et al.*, 1987; Winoto *et al.*, 1989; Redondo *et al.*; 1990
HLA DQ a and/or DQ β	Sullivan *et al.*, 1987
β-Interferon	Goodbourn *et al.*, 1986; Fujita *et al.*, 1987; Goodbourn *et al.*, 1988
Interleukin-2	Greene *et al.*, 1989
Interleukin-2 Receptor	Greene *et al.*, 1989; Lin *et al.*, 1990
MHC Class II 5	Koch *et al.*, 1989
MHC Class II HLA-DRa	Sherman *et al.*, 1989
β-Actin	Kawamoto *et al.*, 1988; Ng *et al.*; 1989
Muscle Creatine Kinase (MCK)	Jaynes *et al.*, 1988; Horlick *et al.*, 1989; Johnson *et al.*, 1989
Prealbumin (Transthyretin)	Costa *et al.*, 1988
Elastase I	Omitz *et al.*, 1987
Metallothionein (MTII)	Karin *et al.*, 1987; Culotta *et al.*, 1989
Collagenase	Pinkert *et al.*, 1987; Angel *et al.*, 1987
Albumin	Pinkert *et al.*, 1987; Tronche *et al.*, 1989, 1990
α-Fetoprotein	Godbout *et al.*, 1988; Campere *et al.*, 1989
t-Globin	Bodine *et al.*, 1987; Perez-Stable *et al.*, 1990
β-Globin	Trudel *et al.*, 1987
c-fos	Cohen *et al.*, 1987
c-HA-*ras*	Triesman, 1986; Deschamps *et al.*, 1985
Insulin	Edlund *et al.*, 1985
Neural Cell Adhesion Molecule (NCAM)	Hirsh *et al.*, 1990

TABLE 1 Promoter and/or Enhancer	
Promoter/Enhancer	References
α_1-Antitrypain	Latimer *et al.*, 1990
H2B (TH2B) Histone	Hwang *et al.*, 1990
Mouse and/or Type I Collagen	Ripe *et al.*, 1989
Glucose-Regulated Proteins (GRP94 and GRP78)	Chang *et al.*, 1989
Rat Growth Hormone	Larsen *et al.*, 1986
Human Serum Amyloid A (SAA)	Edbrooke *et al.*, 1989
Troponin I (TN I)	Yutzey *et al.*, 1989
Platelet-Derived Growth Factor (PDGF)	Pech *et al.*, 1989
Duchenne Muscular Dystrophy	Klamut *et al.*, 1990
SV40	Banerji *et al.*, 1981; Moreau *et al.*, 1981; Sleigh *et al.*, 1985; Firak *et al.*, 1986; Herr *et al.*, 1986; Imbra *et al.*, 1986; Kadesch *et al.*, 1986; Wang *et al.*, 1986; Ondek *et al.*, 1987; Kuhl *et al.*, 1987; Schaffner *et al.*, 1988
Polyoma	Swartzendruber *et al.*, 1975; Vasseur *et al.*, 1980; Katinka *et al.*, 1980, 1981; Tyndell *et al.*, 1981; Dandolo *et al.*, 1983; de Villiers *et al.*, 1984; Hen *et al.*, 1986; Satake *et al.*, 1988; Campbell *et al.*, 1988
Retroviruses	Kriegler *et al.*, 1982, 1983; Levinson *et al.*, 1982; Kriegler *et al.*, 1983, 1984a, b, 1988; Bosze *et al.*, 1986; Miksicek *et al.*, 1986; Celander *et al.*, 1987; Thiesen *et al.*, 1988; Celander *et al.*, 1988; Chol *et al.*, 1988; Reisman *et al.*, 1989
Papilloma Virus	Campo *et al.*, 1983; Lusky *et al.*, 1983; Spandidos and Wilkie, 1983; Spalholz *et al.*, 1985; Lusky *et al.*, 1986; Cripe *et al.*, 1987; Gloss *et al.*, 1987; Hirochika *et al.*, 1987; Stephens *et al.*, 1987
Hepatitis B Virus	Bulla *et al.*, 1986; Jameel *et al.*, 1986; Shaul *et al.*, 1987; Spandau *et al.*, 1988; Vannice *et al.*, 1988
Human Immunodeficiency Virus	Muesing *et al.*, 1987; Hauber *et al.*, 1988; Jakobovits *et al.*, 1988; Feng *et al.*, 1988; Takebe *et al.*, 1988; Rosen *et al.*, 1988; Berkhout *et al.*, 1989; Laspia *et al.*, 1989; Sharp *et al.*, 1989; Braddock *et al.*, 1989

TABLE 1 Promoter and/or Enhancer	
Promoter/Enhancer	References
Cytomegalovirus (CMV)	Weber *et al.*, 1984; Boshart *et al.*, 1985; Foecking *et al.*, 1986
Gibbon Ape Leukemia Virus	Holbrook *et al.*, 1987; Quinn *et al.*, 1989

TABLE 2 Inducible Elements		
Element	Inducer	References
MT II	Phorbol Ester (TFA) Heavy metals	Palmiter *et al.*, 1982; Haslinger *et al.*, 1985; Searle *et al.*, 1985; Stuart *et al.*, 1985; Imagawa *et al.*, 1987, Karin *et al.*, 1987; Angel *et al.*, 1987b; McNeall *et al.*, 1989
MMTV (mouse mammary tumor virus)	Glucocorticoids	Huang *et al.*, 1981; Lee *et al.*, 1981; Majors *et al.*, 1983; Chandler *et al.*, 1983; Lee *et al.*, 1984; Ponta *et al.*, 1985; Sakai *et al.*, 1988
β-Interferon	poly(rI)x poly(rc)	Tavernier *et al.*, 1983
Adenovirus 5 E2	ElA	Imperiale *et al.*, 1984
Collagenase	Phorbol Ester (TPA)	Angel *et al.*, 1987a
Stromelysin	Phorbol Ester (TPA)	Angel *et al.*, 1987b
SV40	Phorbol Ester (TPA)	Angel *et al.*, 1987b
Murine MX Gene	Interferon, Newcastle Disease Virus	Hug *et al.*, 1988
GRP78 Gene	A23187	Resendez *et al.*, 1988
α-2-Macroglobulin	IL-6	Kunz *et al.*, 1989
Vimentin	Serum	Rittling *et al.*, 1989
MHC Class I Gene H-2κb	Interferon	Blanar *et al.*, 1989

TABLE 2 Inducible Elements		
Element	Inducer	References
HSP70	ElA, SV40 Large T Antigen	Taylor *et al.*, 1989, 1990a, 1990b
Proliferin	Phorbol Ester-TPA	Mordacq *et al.*, 1989
Tumor Necrosis Factor	PMA	Hensel *et al.*, 1989
Thyroid Stimulating Hormone α Gene	Thyroid Hormone	Chatterjee *et al.*, 1989

The identity of tissue-specific promoters or elements, as well as assays to characterize their activity, is well known to those of skill in the art. Examples of such regions include the human LIMK2 gene (Nomoto *et al.* 1999), the somatostatin receptor 2 gene (Kraus *et al.*, 1998), murine epididymal retinoic acid-binding gene (Lareyre *et al.*, 1999), human CD4 (Zhao-Emonet *et al.*, 1998), mouse alpha2 (XI) collagen (Tsumaki, *et al.*, 1998), D1A dopamine receptor gene (Lee, *et al.*, 1997), insulin-like growth factor II (Wu *et al.*, 1997), human platelet endothelial cell adhesion molecule-1 (Almendro *et al.*, 1996), and the SM22α promoter.

b. Initiation Signals and Internal Ribosome Binding Sites

A specific initiation signal also may be required for efficient translation of coding sequences. These signals include the ATG initiation codon or adjacent sequences. Exogenous translational control signals, including the ATG initiation codon, may need to be provided. One of ordinary skill in the art would readily be capable of determining this and providing the necessary signals. It is well known that the initiation codon must be "in-frame" with the reading frame of the desired coding sequence to ensure translation of the entire insert. The exogenous translational control signals and initiation codons can be either natural or synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements.

In certain embodiments of the invention, the use of internal ribosome entry sites (IRES) elements are used to create multigene, or polycistronic, messages. IRES elements are able to bypass the ribosome scanning model of 5′– methylated Cap dependent

translation and begin translation at internal sites (Pelletier and Sonenberg, 1988). IRES elements from two members of the picornavirus family (polio and encephalomyocarditis) have been described (Pelletier and Sonenberg, 1988), as well an IRES from a mammalian message (Macejak and Sarnow, 1991). IRES elements can be linked to heterologous open reading frames. Multiple open reading frames can be transcribed together, each separated by an IRES, creating polycistronic messages. By virtue of the IRES element, each open reading frame is accessible to ribosomes for efficient translation. Multiple genes can be efficiently expressed using a single promoter/enhancer to transcribe a single message (see U.S. Patent 5,925,565 and 5,935,819, herein incorporated by reference).

c. Multiple Cloning Sites

Vectors can include a multiple cloning site (MCS), which is a nucleic acid region that contains multiple restriction enzyme sites, any of which can be used in conjunction with standard recombinant technology to digest the vector. (See Carbonelli *et al.,* 1999, Levenson *et al.,* 1998, and Cocea, 1997, incorporated herein by reference.) "Restriction enzyme digestion" refers to catalytic cleavage of a nucleic acid molecule with an enzyme that functions only at specific locations in a nucleic acid molecule. Many of these restriction enzymes are commercially available. Use of such enzymes is widely understood by those of skill in the art. Frequently, a vector is linearized or fragmented using a restriction enzyme that cuts within the MCS to enable exogenous sequences to be ligated to the vector. "Ligation" refers to the process of forming phosphodiester bonds between two nucleic acid fragments, which may or may not be contiguous with each other. Techniques involving restriction enzymes and ligation reactions are well known to those of skill in the art of recombinant technology.

d. Splicing Sites

Most transcribed eukaryotic RNA molecules will undergo RNA splicing to remove introns from the primary transcripts. Vectors containing genomic eukaryotic sequences may require donor and/or acceptor splicing sites to ensure proper processing of the transcript for protein expression. (See Chandler *et al.*, 1997, incorporated herein by reference.)

e. Termination Signals

The vectors or constructs of the present invention will generally comprise at least one termination signal. A "termination signal" or "terminator" is comprised of the DNA sequences involved in specific termination of an RNA transcript by an RNA polymerase. Thus, in certain embodiments a termination signal that ends the production of an RNA transcript is contemplated. A terminator may be necessary *in vivo* to achieve desirable message levels.

In eukaryotic systems, the terminator region may also comprise specific DNA sequences that permit site-specific cleavage of the new transcript so as to expose a polyadenylation site. This signals a specialized endogenous polymerase to add a stretch of about 200 A residues (polyA) to the 3' end of the transcript. RNA molecules modified with this polyA tail appear to more stable and are translated more efficiently. Thus, in other embodiments involving eukaryotes, it is preferred that that terminator comprises a signal for the cleavage of the RNA, and it is more preferred that the terminator signal promotes polyadenylation of the message. The terminator and/or polyadenylation site elements can serve to enhance message levels and/or to minimize read through from the cassette into other sequences.

Terminators contemplated for use in the invention include any known terminator of transcription described herein or known to one of ordinary skill in the art, including but not limited to, for example, the termination sequences of genes, such as for example the bovine growth hormone terminator or viral termination sequences, such as for example the SV40 terminator. In certain embodiments, the termination signal may be a lack of transcribable or translatable sequence, such as due to a sequence truncation.

f. Polyadenylation Signals

In expression, particularly eukaryotic expression, one will typically include a polyadenylation signal to effect proper polyadenylation of the transcript. The nature of the polyadenylation signal is not believed to be crucial to the successful practice of the invention, and/or any such sequence may be employed. Preferred embodiments include the SV40 polyadenylation signal and/or the bovine growth hormone polyadenylation signal, convenient and/or known to function well in various target cells. Polyadenylation may increase the stability of the transcript or may facilitate cytoplasmic transport.

g. Origins of Replication

In order to propagate a vector in a host cell, it may contain one or more origins of replication sites (often termed "ori"), which is a specific nucleic acid sequence at which replication is initiated. Alternatively an autonomously replicating sequence (ARS) can be employed if the host cell is yeast.

h. Selectable and Screenable Markers

In certain embodiments of the invention, cells containing a nucleic acid construct of the present invention may be identified *in vitro* or *in vivo* by including a marker in the expression vector. Such markers would confer an identifiable change to the cell permitting easy identification of cells containing the expression vector. Generally, a selectable marker is one that confers a property that allows for selection. A positive selectable marker is one in which the presence of the marker allows for its selection, while a negative selectable marker is one in which its presence prevents its selection. An example of a positive selectable marker is a drug resistance marker.

Usually the inclusion of a drug selection marker aids in the cloning and identification of transformants, for example, genes that confer resistance to neomycin, puromycin, hygromycin, DHFR, GPT, zeocin and histidinol are useful selectable markers.

In addition to markers conferring a phenotype that allows for the discrimination of transformants based on the implementation of conditions, other types of markers including screenable markers such as GFP, whose basis is colorimetric analysis, are also contemplated. Alternatively, screenable enzymes such as herpes simplex virus thymidine kinase (*tk*), chloramphenicol acetyltransferase (CAT), or luciferase may be utilized. One of skill in the art would also know how to employ immunologic markers, possibly in conjunction with FACS analysis. The marker used is not believed to be important, so long as it is capable of being expressed simultaneously with the nucleic acid encoding a gene product. Further examples of selectable and screenable markers are well known to one of skill in the art.

B. Host Cells

As used herein, the terms "cell," "cell line," and "cell culture" may be used interchangeably. All of these terms also include their progeny, which is any and all subsequent generations. It is understood that all progeny may not be identical due to deliberate or inadvertent mutations. In the context of expressing a heterologous nucleic acid sequence, "host cell" refers to a prokaryotic or eukaryotic cell, and it includes any transformable organisms that is capable of replicating a vector and/or expressing a heterologous gene encoded by a vector. A host cell can, and has been, used as a recipient for vectors. A host cell may be "transfected" or "transformed," which refers to a process by which exogenous nucleic acid, such as a modified protein-encoding sequence, is transferred or introduced into the host cell. A transformed cell includes the primary subject cell and its progeny.

Host cells may be derived from prokaryotes or eukaryotes, including yeast cells, insect cells, and mammalian cells, depending upon whether the desired result is replication of the vector or expression of part or all of the vector-encoded nucleic acid sequences. Numerous cell lines and cultures are available for use as a host cell, and they can be obtained through the American Type Culture Collection (ATCC), which is an organization that serves as an archive for living cultures and genetic materials (www.atcc.org). An appropriate host can be determined by one of skill in the art based on the vector backbone and the desired result. A plasmid or cosmid, for example, can be introduced into a prokaryote host cell for replication of many vectors. Bacterial cells used as host cells for vector replication and/or expression include DH5α, JM109, and KC8, as well as a number of commercially available bacterial hosts such as SURE® Competent Cells and SOLOPACK™ Gold Cells (STRATAGENE®, La Jolla, CA). Alternatively, bacterial cells such as *E. coli* LE392 could be used as host cells for phage viruses. Appropriate yeast cells include *Saccharomyces cerevisiae*, *Saccharomyces pombe*, and *Pichia pastoris*.

Examples of eukaryotic host cells for replication and/or expression of a vector include HeLa, NIH3T3, Jurkat, 293, Cos, CHO, Saos, and PC12. Stem cell lines and other immature cell lines are specifically contemplated as suitable host cells of the invention. Many host cells from various cell types and organisms are available and would

be known to one of skill in the art. Similarly, a viral vector may be used in conjunction with either a eukaryotic or prokaryotic host cell, particularly one that is permissive for replication or expression of the vector.

Some vectors may employ control sequences that allow it to be replicated and/or expressed in both prokaryotic and eukaryotic cells. One of skill in the art would further understand the conditions under which to incubate all of the above described host cells to maintain them and to permit replication of a vector. Also understood and known are techniques and conditions that would allow large-scale production of vectors, as well as production of the nucleic acids encoded by vectors and their cognate polypeptides, proteins, or peptides.

C. Assays of Transgene Expression

Assays may be employed with the instant invention for determination of the relative efficiency of transgene expression. For example, assays may be used to determine the efficacy of deletion mutants of the *s-ship* promoter in directing expression of exogenous proteins. Similarly, one could produce random or site-specific mutants of the *s-ship* promoter of the invention and assay the efficacy of the mutants in the expression of a given transgene. Alternatively, assays could be used to determine the efficacy of the *s-ship* promoter in directing protein expression when used in conjunction with various different enhancers, terminators or other types of elements potentially used in the preparation of transformation constructs.

For mammals, expression assays may comprise a system utilizing cell lines, or alternatively, whole organisms. Additionally, assays of tissue or developmental specific promoters are generally feasible.

The biological sample to be assayed may comprise nucleic acids isolated from the cells of any plant material according to standard methodologies (Sambrook *et al.*, 1989). The nucleic acid may be genomic DNA or fractionated or whole cell RNA. Where RNA is used, it may be desired to convert the RNA to a complementary DNA. In one embodiment of the invention, the RNA is whole cell RNA; in another, it is poly-A RNA. Normally, the nucleic acid is amplified.

Depending on the format, the specific nucleic acid of interest is identified in the sample directly using amplification or with a second, known nucleic acid following

amplification. Next, the identified product is detected. In certain applications, the detection may be performed by visual means (*e.g.*, ethidium bromide staining of a gel). Alternatively, the detection may involve indirect identification of the product *via* chemiluminescence, radioactive scintigraphy of radiolabel or fluorescent label or even *via* a system using electrical or thermal impulse signals (Affymax Technology; Bellus, 1994).

Following detection, one may compare the results seen in a given sample with a statistically significant reference group of non-transformed control cells. Typically, the non-transformed control cells will be of a genetic background similar to the transformed cells. In this way, it is possible to detect differences in the amount or kind of protein detected in various transformed cells.

As indicated, a variety of different assays are contemplated in the screening of cells or animals of the current invention and associated promoters. These techniques may in cases be used to detect for both the presence and expression of the particular genes as well as rearrangements that may have occurred in the gene construct. The techniques include but are not limited to, fluorescent *in situ* hybridization (FISH), direct DNA sequencing, pulsed field gel electrophoresis (PFGE) analysis, Southern or Northern blotting, single-stranded conformation analysis (SSCA), RNAse protection assay, allele-specific oligonucleotide (ASO), dot blot analysis, denaturing gradient gel electrophoresis, RFLP and PCR™-SSCP.

1. Quantitation of Gene Expression with Relative Quantitative RT- PCR™

Reverse transcription (RT) of RNA to cDNA followed by relative quantitative PCR™ (RT-PCR™) can be used to determine the relative concentrations of specific mRNA species, for example, an mRNA whose expression is controlled by an *s-ship* promoter. By determining that the concentration of a specific mRNA species varies, it can be shown that the gene encoding the specific mRNA species is differentially expressed. In this way, a promoters expression profile can be rapidly identified, as can the efficacy with which the promoter directs transgene expression.

In PCR™, the number of molecules of the amplified target DNA increase by a factor approaching two with every cycle of the reaction until some reagent becomes

limiting. Thereafter, the rate of amplification becomes increasingly diminished until there is no increase in the amplified target between cycles. If a graph is plotted in which the cycle number is on the X axis and the log of the concentration of the amplified target DNA is on the Y axis, a curved line of characteristic shape is formed by connecting the plotted points. Beginning with the first cycle, the slope of the line is positive and constant. This is said to be the linear portion of the curve. After a reagent becomes limiting, the slope of the line begins to decrease and eventually becomes zero. At this point the concentration of the amplified target DNA becomes asymptotic to some fixed value. This is said to be the plateau portion of the curve.

The concentration of the target DNA in the linear portion of the PCR™ amplification is directly proportional to the starting concentration of the target before the reaction began. By determining the concentration of the amplified products of the target DNA in PCR™ reactions that have completed the same number of cycles and are in their linear ranges, it is possible to determine the relative concentrations of the specific target sequence in the original DNA mixture. If the DNA mixtures are cDNAs synthesized from RNAs isolated from different tissues or cells, the relative abundances of the specific mRNA from which the target sequence was derived can be determined for the respective tissues or cells. This direct proportionality between the concentration of the PCR™ products and the relative mRNA abundances is only true in the linear range of the PCR™ reaction.

The final concentration of the target DNA in the plateau portion of the curve is determined by the availability of reagents in the reaction mix and is independent of the original concentration of target DNA. Therefore, the first condition that must be met before the relative abundances of a mRNA species can be determined by RT-PCR™ for a collection of RNA populations is that the concentrations of the amplified PCR™ products must be sampled when the PCR™ reactions are in the linear portion of their curves.

The second condition that must be met for an RT-PCR™ study to successfully determine the relative abundances of a particular mRNA species is that relative concentrations of the amplifiable cDNAs must be normalized to some independent standard. The goal of an RT-PCR™ study is to determine the abundance of a particular mRNA species relative to the average abundance of all mRNA species in the sample.

Most protocols for competitive PCR™ utilize internal PCR™ standards that are approximately as abundant as the target. These strategies are effective if the products of the PCR™ amplifications are sampled during their linear phases. If the products are sampled when the reactions are approaching the plateau phase, then the less abundant product becomes relatively over represented. Comparisons of relative abundances made for many different RNA samples, such as is the case when examining RNA samples for differential expression, become distorted in such a way as to make differences in relative abundances of RNAs appear less than they actually are. This is not a significant problem if the internal standard is much more abundant than the target. If the internal standard is more abundant than the target, then direct linear comparisons can be made between RNA samples.

The above discussion describes theoretical considerations for an RT-PCR™ assay for plant tissue. The problems inherent in plant tissue samples are that they are of variable quantity (making normalization problematic), and that they are of variable quality (necessitating the co-amplification of a reliable internal control, preferably of larger size than the target). Both of these problems are overcome if the RT-PCR™ is performed as a relative quantitative RT-PCR™ with an internal standard in which the internal standard is an amplifiable cDNA fragment that is larger than the target cDNA fragment and in which the abundance of the mRNA encoding the internal standard is roughly 5-100 fold higher than the mRNA encoding the target. This assay measures relative abundance, not absolute abundance of the respective mRNA species.

Other studies may be performed using a more conventional relative quantitative RT-PCR™ assay with an external standard protocol. These assays sample the PCR™ products in the linear portion of their amplification curves. The number of PCR™ cycles that are optimal for sampling must be empirically determined for each target cDNA fragment. In addition, the reverse transcriptase products of each RNA population isolated from the various tissue samples must be carefully normalized for equal concentrations of amplifiable cDNAs. This consideration is very important since the assay measures absolute mRNA abundance. Absolute mRNA abundance can be used as a measure of differential gene expression only in normalized samples. While empirical determination of the linear range of the amplification curve and normalization of cDNA preparations

are tedious and time consuming processes, the resulting RT-PCR™ assays can be superior to those derived from the relative quantitative RT-PCR™ assay with an internal standard.

One reason for this advantage is that without the internal standard/competitor, all of the reagents can be converted into a single PCR™ product in the linear range of the amplification curve, thus increasing the sensitivity of the assay. Another reason is that with only one PCR™ product, display of the product on an electrophoretic gel or another display method becomes less complex, has less background and is easier to interpret.

2. Marker Gene Expression

Marker genes represent an efficient means for assaying the expression of transgenes. Using, for example, a selectable marker gene, one could quantitatively determine the expression levels in the cell using a construct comprising the selectable marker coding region operably linked to the promoter to be assayed, *e.g.*, an *s-ship* promoter. Alternatively, particular cell types could be exposed to a selective agent and the relative resistance provided in these cells quantified, thereby providing an estimate of the tissue specific expression of the promoter.

Screenable markers constitute another efficient means for quantifying the expression of a given transgene. Potentially any screenable marker could be expressed and the marker gene product quantified, thereby providing an estimate of the efficiency with which the promoter directs expression of the transgene. Quantification can readily be carried out using either visual means, or, for example, a photon counting device.

A preferred screenable marker gene assay for use with the current invention include the use of the screenable marker gene β-galactosidase (β-gal), luciferase, or green fluorescent protein (GFP).

3. Purification and Assays of Proteins

One means for determining the efficiency with which a particular transgene is expressed is to purify and quantify a polypeptide expressed by the transgene. Protein purification techniques are well known to those of skill in the art. These techniques involve, at one level, the crude fractionation of the cellular milieu to polypeptide and non-polypeptide fractions. Having separated the polypeptide from other proteins, the

polypeptide of interest may be further purified using chromatographic and electrophoretic techniques to achieve partial or complete purification (or purification to homogeneity). Analytical methods particularly suited to the preparation of a pure peptide are ion-exchange chromatography, exclusion chromatography; polyacrylamide gel electrophoresis; and isoelectric focusing. A particularly efficient method of purifying peptides is fast protein liquid chromatography or even HPLC.

Various techniques suitable for use in protein purification will be well known to those of skill in the art. These include, for example, precipitation with ammonium sulphate, PEG, antibodies and the like or by heat denaturation, followed by centrifugation; chromatography steps such as ion exchange, gel filtration, reverse phase, hydroxylapatite and affinity chromatography; isoelectric focusing; gel electrophoresis; and combinations of such and other techniques. As is generally known in the art, it is believed that the order of conducting the various purification steps may be changed, or that certain steps may be omitted, and still result in a suitable method for the preparation of a substantially purified protein or peptide.

There is no general requirement that the protein or peptide being assayed always be provided in their most purified state. Indeed, it is contemplated that less substantially purified products will have utility in certain embodiments. Partial purification may be accomplished by using fewer purification steps in combination, or by utilizing different forms of the same general purification scheme. For example, it is appreciated that a cation-exchange column chromatography performed utilizing an HPLC apparatus will generally result in a greater "-fold" purification than the same technique utilizing a low pressure chromatography system. Methods exhibiting a lower degree of relative purification may have advantages in total recovery of protein product, or in maintaining the activity of an expressed protein.

It is known that the migration of a polypeptide can vary, sometimes significantly, with different conditions of SDS/PAGE (Capaldi *et al.,* 1977). It will therefore be appreciated that under differing electrophoresis conditions, the apparent molecular weights of purified or partially purified expression products may vary.

High Performance Liquid Chromatography (HPLC) is characterized by a very rapid separation with extraordinary resolution of peaks. This is achieved by the use of

very fine particles and high pressure to maintain an adequate flow rate. Separation can be accomplished in a matter of minutes, or at most an hour. Moreover, only a very small volume of the sample is needed because the particles are so small and close-packed that the void volume is a very small fraction of the bed volume. Also, the concentration of the sample need not be very great because the bands are so narrow that there is very little dilution of the sample.

Gel chromatography, or molecular sieve chromatography, is a special type of partition chromatography that is based on molecular size. The theory behind gel chromatography is that the column, which is prepared with tiny particles of an inert substance that contain small pores, separates larger molecules from smaller molecules as they pass through or around the pores, depending on their size. As long as the material of which the particles are made does not adsorb the molecules, the sole factor determining rate of flow is the size. Hence, molecules are eluted from the column in decreasing size, so long as the shape is relatively constant. Gel chromatography is unsurpassed for separating molecules of different size because separation is independent of all other factors such as pH, ionic strength, temperature, *etc.* There also is virtually no adsorption, less zone spreading and the elution volume is related in a simple matter to molecular weight.

Affinity Chromatography is a chromatographic procedure that relies on the specific affinity between a substance to be isolated and a molecule that it can specifically bind to. This is a receptor-ligand type interaction. The column material is synthesized by covalently coupling one of the binding partners to an insoluble matrix. The column material is then able to specifically adsorb the substance from the solution. Elution occurs by changing the conditions to those in which binding will not occur (alter pH, ionic strength, temperature, *etc.*).

A particular type of affinity chromatography useful in the purification of carbohydrate containing compounds is lectin affinity chromatography. Lectins are a class of substances that bind to a variety of polysaccharides and glycoproteins.

The matrix should be a substance that itself does not adsorb molecules to any significant extent and that has a broad range of chemical, physical and thermal stability. The ligand should be coupled in such a way as to not affect its binding properties. The

ligand should also provide relatively tight binding. And it should be possible to elute the substance without destroying the sample or the ligand. One of the most common forms of affinity chromatography is immunoaffinity chromatography. The generation of antibodies that would be suitable for use in accord with the present invention is well known to those of skill in the art.

D. Methods of Gene Transfer

Suitable methods for nucleic acid delivery to effect expression of compositions of the present invention are believed to include virtually any method by which a nucleic acid (*e.g.*, DNA, including viral and nonviral vectors) can be introduced into an organelle, a cell, a tissue or an organism, as described herein or as would be known to one of ordinary skill in the art. Such methods include, but are not limited to, direct delivery of DNA such as by injection (U.S. Patents 5,994,624, 5,981,274, 5,945,100, 5,780,448, 5,736,524, 5,702,932, 5,656,610, 5,589,466 and 5,580,859, each incorporated herein by reference), including microinjection (Harlan and Weintraub, 1985; U.S. Patent 5,789,215, incorporated herein by reference); by electroporation (U.S. Patent No. 5,384,253, incorporated herein by reference); by calcium phosphate precipitation (Graham and Van Der Eb, 1973; Chen and Okayama, 1987; Rippe *et al.*, 1990); by using DEAE-dextran followed by polyethylene glycol (Gopal, 1985); by direct sonic loading (Fechheimer *et al.*, 1987); by liposome mediated transfection (Nicolau and Sene, 1982; Fraley *et al.*, 1979; Nicolau *et al.*, 1987; Wong *et al.*, 1980; Kaneda *et al.*, 1989; Kato *et al.*, 1991); by microprojectile bombardment (PCT Application Nos. WO 94/09699 and 95/06128; U.S. Patents 5,610,042; 5,322,783 5,563,055, 5,550,318, 5,538,877 and 5,538,880, and each incorporated herein by reference); by agitation with silicon carbide fibers (Kaeppler *et al.*, 1990; U.S. Patents 5,302,523 and 5,464,765, each incorporated herein by reference); by desiccation/inhibition-mediated DNA uptake (Potrykus *et al.*, 1985). Through the application of techniques such as these, organelle(s), cell(s), tissue(s) or organism(s) may be stably or transiently transformed.

E. Transgenic and Knockout Animals

1. Transgenic Animals

It is further contemplated that transgenic animals are part of the present invention. A transgenic animal of the present invention may involve an animal in which an *s-ship*

promoter drives the expression of a transgene. The transgene can be expressed temporally or spatially in a manner different than or the same as a non-transgenic animal. The transgene may also be heterologous with respect to the host cell or organism, such as, for example, the luciferase gene in a mammalian cell. Moreover, it is contemplated that the transgene may be expressed in a different tissue type or in a different amount or at a different time than the endogenously expressed version of the transgene.

In a general aspect, a transgenic animal is produced by the integration of a given transgene into the genome in a manner that permits the expression of the transgene, or by disrupting the wild-type gene, leading to a knockout of the wild-type gene. Methods for producing transgenic animals are generally described by Wagner and Hoppe (U.S. Patent No. 4,873,191; which is incorporated herein by reference), Brinster *et al.* (1985; which is incorporated herein by reference in its entirety) and in "Manipulating the Mouse Embryo; A Laboratory Manual" 2nd edition (eds., Hogan, Beddington, Costantimi and Long, Cold Spring Harbor Laboratory Press, 1994; which is incorporated herein by reference in its entirety).

U.S. Patent 5,639,457 is also incorporated herein by reference to supplement the present teaching regarding transgenic pig and rabbit production. U.S. Patents 5,175,384; 5,175,385; 5,530,179, 5,625,125, 5,612,486 and 5,565,186 are also each incorporated herein by reference to similarly supplement the present teaching regarding transgenic mouse and rat production. Transgenic animals may be crossed with other transgenic animals or knockout animals to evaluate phenotype based on compound alterations in the genome.

2. Knockout Animals or Cells

The generation of an animal model lacking s-ship or a particular nucleic acid (encoding an RNA that is translated or not) is contemplated as part of the present invention to understand further stem cell function. This strategy could also be implemented in cell culture as well.

The lack of activity as a result of the knockout may provoke various types of pathophysiological disturbances in a knockout animal or cell. This can be used to characterize the role or function of a particular gene product at a particular time in development or in a particular cell type. Use of the *s-ship* promoter can be used to drive

the expression of the knockout gene such that only certain cells, for example stem cells, may be affected. One method of inhibiting the endogenous expression of a particular gene in an animal is to disrupt the gene in germline cells and produce offspring from these cells. This method is generally known as knockout technology. U.S. Patent No. 5,616,491, incorporated herein by reference in its entirety, generally describes the techniques involved in the preparation of knockout mice, and in particular describes mice having a suppressed level of expression of the gene encoding CD28 on T cells, and mice wherein the expression of the gene encoding CD45 is suppressed on B cells. Pfeffer *et al.* (1993) describe mice in which the gene encoding the tumor necrosis factor receptor p55 has been suppressed. The mice showed a decreased response to tumor necrosis factor signaling. Fung-Leung *et al.* (1991a; 1991b) describe knockout mice lacking expression of the gene encoding CD8. These mice were found to have a decreased level of cytotoxic T cell response to various antigens and to certain viral pathogens such as lymphocytic choriomeningitis virus.

The term "knockout" refers to a partial or complete suppression of the expression of at least a portion of a protein encoded by an endogenous DNA sequence in a cell. The term "knockout construct" refers to a nucleic acid sequence that is designed to decrease or suppress expression of a protein encoded by endogenous DNA sequences in a cell. The nucleic acid sequence used as the knockout construct is typically comprised of: (1) DNA from some portion of the gene (exon sequence, intron sequence, and/or promoter sequence) to be suppressed, in conjunction with all or part of the *s-ship* promoter; and (2) a marker sequence used to detect the presence of the knockout construct in the cell. The knockout construct is inserted into a cell, and integrates with the genomic DNA of the cell in such a position so as to prevent or interrupt transcription of the native DNA sequence. Such insertion usually occurs by homologous recombination (*i.e.*, regions of the knockout construct that are homologous to endogenous DNA sequences hybridize to each other when the knockout construct is inserted into the cell and recombine so that the knockout construct is incorporated into the corresponding position of the endogenous DNA).

The knockout construct nucleic acid sequence may comprise 1) a full or partial sequence of one or more exons and/or introns of the gene to be suppressed, 2) a full or

partial promoter sequence of the gene to be suppressed, or 3) combinations thereof. Typically, the knockout construct is inserted into an embryonic stem cell (ES cell) and is integrated into the ES cell genomic DNA, usually by the process of homologous recombination. This ES cell is then injected into, and integrates with, the developing embryo.

The phenotype of a mouse heterozygous for the knockout may lend clues as to the function and importance of that gene or sequence, as well as contribute an understanding about its physiological relevance, particularly with respect to disease states. Animals completely lacking the targeted gene (homozygous null) may provide additional information. Mice lacking the targeted gene may not be viable, which itself is indicative of the importance of that gene. Should such mice be viable (heterozygous or homozygous nulls), they may be crossed with other transgenic or knockout mice. Furthermore, knock-out mice having any phenotype that resembles a disease state may be used to screen or test therapeutic drugs that slow, modify, or cure conditions. As is known to the skilled artisan, a conditional knockout, wherein the gene is disrupted under certain conditions, is frequently used.

3. Conditional Transgenic and Knockdown Animals and Cells

The present invention further contemplates conditional transgenic or knockdown animals (or cells in culture), such as those produced using recombination methods. Bacteriophage P1 Cre recombinase and flp recombinase from yeast plasmids are two non-limiting examples of site-specific DNA recombinase enzymes which cleave DNA at specific target sites (lox P sites for cre recombinase and frt sites for flp recombinase) and catalyze a ligation of this DNA to a second cleaved site. A large number of suitable alternative site-specific recombinases have been described, and their genes can be used in accordance with the method of the present invention. Such recombinases include the Int recombinase of bacteriophage λ (with or without Xis) (Weisberg *et. al.*, 1983), herein incorporated by reference); TpnI and the β-lactamase transposons (Mercier *et al.*, 1990); the Tn3 resolvase (Flanagan and Fennewald, 1989; Stark *et al.*, 1989); the yeast recombinases (Matsuzaki *et al.*, 1990); the B. subtilis SpoIVC recombinase (Sato *et al.*, 1990); the Flp recombinase (Schwartz and Sadowski, 1989; Parsons *et al.*, 1990; Golic and Lindquist, 1989; Amin *et al.*, 1990); the Hin recombinase (Glasgow *et al.*, 1989);

immunoglobulin recombinases (Malynn *et al.*, 1988); and the Cin recombinase (Haffter and Bickle, 1988; Hubner *et al.*, 1989), all herein incorporated by reference. Such systems are discussed (Echols, 1990; de Villartay, 1988; Craig, 1988; Poyart-Salmeron *et al.*, 1989; Hunger-Bertling *et al.*, 1990; and Cregg and Madden, 1989), all herein incorporated by reference.

Of particular interest in the present invention is the Cre recombinase. Cre has been purified to homogeneity, and its reaction with the loxP site has been extensively characterized (Abremski and Hess, 1984), herein incorporated by reference). Cre protein has a molecular weight of 35,000 and can be obtained commercially from New England Nuclear/DuPont. The cre gene (which encodes the Cre protein) has been cloned and expressed (Abremski *et al.*, 1983), herein incorporated by reference). The Cre protein mediates recombination between two loxP sequences (Sternberg *et al.*, 1981), which may be present on the same or different DNA molecule. Because the internal spacer sequence of the loxP site is asymmetrical, two loxP sites can exhibit directionality relative to one another (Hoess and Abremski, 1984). Thus, when two sites on the same DNA molecule are in a directly repeated orientation, Cre will excise the DNA between the sites (Abremski *et al.*, 1983). However, if the sites are inverted with respect to each other, the DNA between them is not excised after recombination but is simply inverted. Thus, a circular DNA molecule having two loxP sites in direct orientation will recombine to produce two smaller circles, whereas circular molecules having two loxP sites in an inverted orientation simply invert the DNA sequences flanked by the loxP sites. In addition, recombinase action can result in reciprocal exchange of regions distal to the target site when targets are present on separate DNA molecules.

Recombinases have important application for characterizing gene function in knockout models. When the constructs described herein are used to disrupt limulus clotting factor protease-like genes, a fusion transcript can be produced when insertion of the positive selection marker occurs downstream (3') of the translation initiation site of the limulus clotting factor protease-like gene. The fusion transcript could result in some level of protein expression with unknown consequence. It has been suggested that insertion of a positive selection marker gene can affect the expression of nearby genes. These effects may make it difficult to determine gene function after a knockout event

since one could not discern whether a given phenotype is associated with the inactivation of a gene, or the transcription of nearby genes. Both potential problems are solved by exploiting recombinase activity. When the positive selection marker is flanked by recombinase sites in the same orientation, the addition of the corresponding recombinase will result in the removal of the positive selection marker. In this way, effects caused by the positive selection marker or expression of fusion transcripts are avoided.

III. Proteinaceous Compositions

In certain embodiments, the present invention concerns novel compositions comprising at least one proteinaceous molecule, such as s-SHIP1, SHIP1, or a modulator of an *s-ship1* promoter. As used herein, a "proteinaceous molecule," "proteinaceous composition," "proteinaceous compound," "proteinaceous chain" or "proteinaceous material" generally refers, but is not limited to, a protein of greater than about 200 amino acids or the full length endogenous sequence translated from a gene; a polypeptide of greater than about 100 amino acids; and/or a peptide of from about 3 to about 100 amino acids. All the "proteinaceous" terms described above may be used interchangeably herein.

In certain embodiments the size of the at least one proteinaceous molecule may comprise, but is not limited to, about 5, about 6, about 7, about 8, about 9, about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19, about 20, about 21, about 22, about 23, about 24, about 25, about 26, about 27, about 28, about 29, about 30, about 31, about 32, about 33, about 34, about 35, about 36, about 37, about 38, about 39, about 40, about 41, about 42, about 43, about 44, about 45, about 46, about 47, about 48, about 49, about 50, about 51, about 52, about 53, about 54, about 55, about 56, about 57, about 58, about 59, about 60, about 61, about 62, about 63, about 64, about 65, about 66, about 67, about 68, about 69, about 70, about 71, about 72, about 73, about 74, about 75, about 76, about 77, about 78, about 79, about 80, about 81, about 82, about 83, about 84, about 85, about 86, about 87, about 88, about 89, about 90, about 91, about 92, about 93, about 94, about 95, about 96, about 97, about 98, about 99, about 100, about 110, about 120, about 130, about 140, about 150, about 160, about 170, about 180, about 190, about 200, about 210, about 220, about 230, about 240, about 250, about 275, about 300, about 325, about 350, about 375, about 400, about 425, about 450, about 475,

about 500, about 525, about 550, about 575, about 600, about 625, about 650, about 675, about 700, about 725, about 750, about 775, about 800, about 825, about 850, about 875, about 900, about 925, about 950, about 975, about 1000, about 1100, about 1200, about 1300, about 1400, about 1500, about 1750, about 2000, about 2250, about 2500 or greater amino molecule residues, and any range derivable therein.

As used herein, an "amino molecule" refers to any amino acid, amino acid derivative or amino acid mimic as would be known to one of ordinary skill in the art. In certain embodiments, the residues of the proteinaceous molecule are sequential, without any non-amino molecule interrupting the sequence of amino molecule residues. In other embodiments, the sequence may comprise one or more non-amino molecule moieties. In particular embodiments, the sequence of residues of the proteinaceous molecule may be interrupted by one or more non-amino molecule moieties.

Accordingly, the term "proteinaceous composition" encompasses amino molecule sequences comprising at least one of the 20 common amino acids in naturally synthesized proteins, or at least one modified or unusual amino acid.

Proteinaceous compositions may be made by any technique known to those of skill in the art, including the expression of proteins, polypeptides or peptides through standard molecular biological techniques, the isolation of proteinaceous compounds from natural sources, or the chemical synthesis of proteinaceous materials. The nucleotide and protein, polypeptide and peptide sequences for various genes have been previously disclosed, and may be found at computerized databases known to those of ordinary skill in the art. One such database is the National Center for Biotechnology Information's Genbank and GenPept databases (http://www.ncbi.nlm.nih.gov/). The coding regions for these known genes may be amplified and/or expressed using the techniques disclosed herein or as would be know to those of ordinary skill in the art. Alternatively, various commercial preparations of proteins, polypeptides and peptides are known to those of skill in the art.

In certain embodiments a proteinaceous compound may be purified. Generally, "purified" will refer to a specific or protein, polypeptide, or peptide composition that has been subjected to fractionation to remove various other proteins, polypeptides, or peptides, and which composition substantially retains its activity, as may be assessed, for

example, by the protein assays, as would be known to one of ordinary skill in the art for the specific or desired protein, polypeptide or peptide.

It is contemplated that virtually any protein, polypeptide or peptide containing component may be used in the compositions and methods disclosed herein. However, it is preferred that the proteinaceous material is biocompatible.

IV. Therapeutic Applications

The invention is widely applicable to a variety of situations where it is desirable to be able to regulate the level of gene expression, such as by turning gene expression "on" and "off", in a rapid, efficient and controlled manner without causing pleiotropic effects or cytotoxicity. The invention may be particularly useful for gene therapy purposes in humans, in treatments for either genetic or acquired diseases. The general approach of gene therapy involves the introduction of one or more nucleic acid molecules into cells such that one or more gene products encoded by the introduced genetic material are produced in the cells to restore or enhance a functional activity. For reviews on gene therapy approaches Anderson, *et al.* (1992; Miller *et al.* (1992); Friedmann *et al.* (1989); and Cournoyer *et al.* (1990). However, current gene therapy vectors typically utilize constitutive regulatory elements which are responsive to endogenous transcriptions factors. These vector systems do not allow for the ability to modulate the level of gene expression in a subject. In contrast, the regulatory system of the invention provides this ability.

To use the system of the invention for gene therapy purposes, at least one DNA molecule is introduced into cells of a subject in need of gene therapy (*e.g.*, a human subject suffering from a genetic or acquired disease) to modify the cells. The cells are modified to comprise: 1) nucleic acid encoding an inducible regulator of the invention in a form suitable for expression of the inducible regulator in the host cells; and 2) an siRNA (*e.g.*, for therapeutic purposes) operatively linked to a tissue-specific promoter such as an *s-ship1* promoter. A single DNA molecule encoding components of the regulatory system of the invention can be used, or alternatively, separate DNA molecules encoding each component can be used. The cells of the subject can be modified *ex vivo* and then introduced into the subject or the cells can be directly modified *in vivo* by conventional techniques for introducing nucleic acid into cells. Thus, the regulatory

system of the invention offers the advantage over constitutive regulatory systems of allowing for modulation of the level of gene expression depending upon the requirements of the therapeutic situation.

Genes of particular interest to be knocked down or knocked out in cells of a subject for treatment of genetic or acquired diseases include those encoding a deleterious gene product, such as an abnormal protein. Examples of non-limiting specific diseases include anemia, blood-related cancers, Parkinson's disease, and diabetes.

The present invention can be applied to develop autologous or allogeneic cell lines for therapeutical purposes. For example, gene therapy applications of particular interest in cell and/or organ transplantation are utilized with the present invention. In exemplary embodiments, downregulation of transplantation antigens (such as, for example, by downregulation of beta2-microglobulin expression *via* siRNA) allows for transplantation of allogeneic cells while minimizing the risk of rejection by the patient's immune system. The present invention would allow for a switch off of the RNAi in case of adverse effects (*e.g.* uncontrollable replication of the transplanted cells).

Cells types that can be subjected to the present invention include hematopoietic stem cells, myoblasts, hepatocytes, lymphocytes, airway epithelium, skin epithelium, islets, dopaminergic neurons, keratinocytes, and so forth. For further descriptions of cell types, genes and methods for gene therapy see *e.g.*, Wilson *et al.* (1988); Armentano *et al.* (1990); Wolff *et al.* (1990); Chowdhury *et al.* (1991); Ferry *et al.* (1991); Wilson *et al.* (1992); Quantin *et al.* (1992); Dai *et al.* (1992); van Beusechem *et al.* (1992); Rosenfeld *et al.* (1992); Kay *et al.* (1992); Cristiano *et al.* (1993); Hwu *et al.* (1993); and Herz and Gerard (1993).

In particular embodiments of the present invention, there is a method of treating any disease condition amenable to treatment with an *s-ship* promoter. In specific embodiments, the method comprises preparing a polynucleotide construct having a region encoding a therapeutic or diagnostic (marker) gene that is operably linked to an an *s-ship* promoter, wherein the gene encoded by the construct is for the treatment of the disease condition.

A. Pharmaceutical Formulations, Delivery, and Treatment Regimens

In an embodiment of the present invention, methods of treatment are contemplated. An effective amount of the pharmaceutical composition, generally, is defined as that amount sufficient to detectably and repeatedly to ameliorate, reduce, minimize or limit the extent of the disease or its symptoms. More rigorous definitions may apply, including elimination, eradication or cure of disease.

The routes of administration will vary, naturally, with the location and nature of the lesion, and include, *e.g.*, intradermal, transdermal, parenteral, intravenous, intramuscular, intranasal, subcutaneous, percutaneous, intratracheal, intraperitoneal, intratumoral, perfusion, lavage, direct injection, and oral administration and formulation.

Solutions of the active compounds as free base or pharmacologically acceptable salts may be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions may also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations contain a preservative to prevent the growth of microorganisms. The pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions (U.S. Patent 5,466,468, specifically incorporated herein by reference in its entirety). In all cases the form must be sterile and must be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (*e.g.*, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and/or vegetable oils. Proper fluidity may be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. The prevention of the action of microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in

the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

For parenteral administration in an aqueous solution, for example, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intramuscular, subcutaneous, intratumoral and intraperitoneal administration. In this connection, sterile aqueous media that can be employed will be known to those of skill in the art in light of the present disclosure. For example, one dosage may be dissolved in 1 ml of isotonic NaCl solution and either added to 1000 ml of hypodermoclysis fluid or injected at the proposed site of infusion, (see for example, "Remington's Pharmaceutical Sciences" 15th Edition, pages 1035-1038 and 1570-1580). Some variation in dosage will necessarily occur depending on the condition of the subject being treated. The person responsible for administration will, in any event, determine the appropriate dose for the individual subject. Moreover, for human administration, preparations should meet sterility, pyrogenicity, general safety and purity standards as required by FDA Office of Biologics standards.

Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized active ingredients into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum-drying and freeze-drying techniques which yield a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

The compositions disclosed herein may be formulated in a neutral or salt form. Pharmaceutically-acceptable salts, include the acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium,

calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like. Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective. The formulations are easily administered in a variety of dosage forms such as injectable solutions, drug release capsules and the like.

As used herein, "carrier" includes any and all solvents, dispersion media, vehicles, coatings, diluents, antibacterial and antifungal agents, isotonic and absorption delaying agents, buffers, carrier solutions, suspensions, colloids, and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.

The phrase "pharmaceutically-acceptable" or "pharmacologically-acceptable" refers to molecular entities and compositions that do not produce an allergic or similar untoward reaction when administered to a human. The preparation of an aqueous composition that contains a protein as an active ingredient is well understood in the art. Typically, such compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection can also be prepared.

B. Combination Treatments

The compounds and methods of the present invention may be used in the context of traditional therapies. In order to increase the effectiveness of a treatment with the compositions of the present invention, it may be desirable to combine these compositions with other agents effective in the treatment of those diseases and conditions. For example, the treatment of a cancer may be implemented with therapeutic compounds of the present invention and other anti-cancer therapies, such as anti-cancer agents or surgery. Likewise, the treatment of a vascular disease or condition may involve both the present invention and conventional vascular agents or therapies.

Various combinations may be employed; for example, a host cell of the present invention is "A" and the secondary anti-cancer agent/therapy is "B":

A/B/A B/A/B B/B/A A/A/B A/B/B B/A/A A/B/B/B B/A/B/B

B/B/B/A B/B/A/B A/A/B/B A/B/A/B A/B/B/A B/B/A/A

B/A/B/A B/A/A/B A/A/A/B B/A/A/A A/B/A/A A/A/B/A

Administration of the therapeutic expression constructs of the present invention to a patient will follow general protocols for the administration of that particular secondary therapy, taking into account the toxicity, if any, of the treatment. It is expected that the treatment cycles would be repeated as necessary. It also is contemplated that various standard therapies, as well as surgical intervention, may be applied in combination with the described therapy.

V. EXAMPLES

The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventor to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

EXAMPLE 1:

Materials and Methods

Cell growth and transfection conditions

NIH3T3 cells, originally obtained from the American Type Culture Collection (ATCC, Rockville, Maryland), were grown in DMEM with 10% fetal bovine serum. The D3 embryonic stem (ES) cell line was obtained from Dr. Tasuku Honjo (Nakano *et al.*, 1994) and grown in high glucose DMEM (GIBCO/Invitrogen Corp., #11965-092) supplemented with 2 mM L-glutamine, 1 mM sodium pyruvate, 0.1 mM nonessential

amino acids, 0.15 mM monothioglycerol (Sigma, M7522), and 15% fetal bovine serum (pre-tested for ES cell growth (HyClone Labs, Inc.)). D3 ES cells were routinely grown on a LIF-producing feeder layer of mitomycin C-treated (Nagy *et al.*, 2003) SNL cells, obtained from Phil Soriano (FHCRC). The SNL cells are G418-resistant. Usually, one passage before flow cytometry, ES cell were transferred to gelatin(Sigma)-coated plates without a feeder layer and with LIF (ESGRO) added to the medium (1000 units/ml).

DNA was transfected into D3 ES cells by electroportion essentially as described by Nagy *et al.*, (2003). ES cells were suspended in PBS (Ca^{2+} and Mg^{2+}- free) at 1 x 10^6 cells/ml and 0.8 ml of the cell suspension placed in a 0.4-cm-wide electrode-gap sterile cuvette (BIO-RAD). Plasmid DNA (20 μg), linearized by overnight digestion with *Afl* II and Qiagen-purified, was added and mixed. Two pulses (instead of one as recommended) of current were applied to the cells in the cuvette employing settings of 500mF, and 230V on a BIO-RAD Gene-Pulser™ with Capacitance Extender. After 5 min on ice, the viscous solution was transferred to a 10-cm culture dish containing mitomycin C-treated SNL cells. After 24 hr, G418 selection was begun using 280 μg/ml active G418. Cells were passed after 10-14 days onto gelatin-coated plates (no feeder cells) in LIF containing medium with G418. Flow cytometry was performed 3-4 days later.

Afl II-linearized plasmid DNA (10 μg) was introduced into NIH3T3 cells by transfection using Superfect reagent (Qiagen) as recommended by the manufacturer. G418 selection was begun 24 hr after transfection using 400 μg/ml G418. Cells were passaged twice in G418 before flow cytometry. Regardless of the electroporation into ES cells or transfection into the NIH3T3 cells, abundant G418 resistant colonies were obtained for each cell type.

Two positive control GFP-expression plasmids were used for both NIH3T3 cells and the D3 ES cells to be sure the transfection/electroporation steps were functional and that GFP expression occurred in each experiment. These positive controls also helped set the gates for analyses of GFP -expressing cells. These two plasmids were the pIRES2-GFP empty plasmid (BD Biosciences Clontech) and pIRES2-GFP containing an insert encoding the Capn5 gene. Both plasmids expressed equally well in each cell type, and the

empty pIRES2-GFP vector always expressed higher levels of GFP than the one containing the insert.

Immunoblotting analysis for SHIP proteins

The techniques for cell extraction, electrophoresis, and immunoblotting have been described previously (Liu *et al.*, 2110). Equal amounts of protein extracts from each cell type were loaded for gel electrophoresis. SHIP proteins were detected using monoclonal antibody P2C6 at a 1:1000 dilution (Lucas and Rohrschneider, 1999).

Flow cytometry

Cells were examined for GFP expression on a Caliber II bench-top analyzer. Cytometer setting were established using positive FDC-P1 cells expressing GFP from a retroviral vector and negative cells, not transfected, or transfected with an empty plasmid. At least 10^4 cells were analyzed for each plasmid transfected, and two independent transfections were examined. Both transfections gave similar results, and the results of one experiment are shown.

Construction of promoter-less GFP-expression constructs for analysis of s-SHIP intron-5 promoter activity

A 7.6-kb DNA *Sac* I-*Sac* I fragment from a Lambda 129Sv mouse genomic clone (Wolf *et al.*, 2000, NCBI accession #AF235499, hereby incorporated by reference) was used for initial examination of potential tissue-specific promoter activity. This region contained almost all of intron-5, the 88 bp of exon-6, and 1271 bp extending into intron-6. This 7.6-kb segment was cloned into pBluescript KS (Stratagene), and sub-segments of the region were obtained with the restriction sites shown in FIG. 2. These sub-segments were cloned into a promoter-less GFP-expression construct.

The promoter-less GFP-expression construct was made from the pEGFP-1 plasmid (BD Biosciences Clontech) by modifications of the MCS (multiple cloning site), incorporating additional synthesized cloning sites (EcoRI-AccI(up)-BssHII-NheI-PstI) for insertion of the sub-fragments from the 7.6 kb intron-5 clone. Both AccI and BssHII recognize multiple sequences and the nucleotide sequence in the synthesized DNA corresponded to AccI site at nucleotide 2776 of the 7.6-kb region, and the 5' BssHII site

of the pBluescript plasmid, respectively. In addition, prior to incorporation of the extended MCS, the SV40 early and late introns from pCMVβ were inserted at the 3' end of the MCS between the KpnI and AgeI sites. Two intron cassettes were used: one containing only the splice acceptor site from the long intron, and a second containing both early and late introns. The former was used only for inserts (*e.g.*, the 7.6-kb and 4.2-kb inserts) containing an intact exon 6 with its splice donor site. The two final plasmids each containing the extended MCS and either the late SV40 intron only (pEGFP2-SD3-1), or both SV40 introns (pEGFP2-SD1-2), were sequenced through the inserted intron region and one of each with correct sequence selected for inserting the 7.6-kb clone and sub-regions.

The longest promoter construct contained the entire 7.6-kb putative *s-SHIP* promoter region, and was excised from the pBluescript plasmid with BssHII for insertion into the MCS of the pEGFP-SD3-1 plasmid. The 6.3-kb fragment was obtained with a partial PstI digestion and complete BssHII digestion. The 4.4-kb and 4.2-kb fragments were from derived from PstI and AccI digestions, respectively. The 1.9-kb segment was obtained from digestion of the 4.4 kb fragment with NheI. The smallest 0.96 kb region was produced by deleting a region of the pBluescript 7.6 kb clone from the SwaI site 960 nucleotides 5' of exon 6, to the FbaI site 22 nucleotides from the 5' end of the 7.6 kb clone. After ligation, the fragment from the 5' BssHII site to the PstI site was excised. Each fragment was inserted into their respective restriction sites of the extended MCS. Restriction analysis of each purified plasmid confirmed the correct insert in the correct orientation, and all cloning junctions were sequenced to confirm proper ligation. Each plasmid was linearized with AflII , and Qiagen purified from agarose gels before electroporation or transfection.

Construction of the 11.5kb- and 6.2kb-GFP *s-SHIP* promoter transgenes

The 11.5kb-GFP transgenic construct was prepared from two separate plasmids containing the two halves of the proposed *s-SHIP* promoter region, plus an 833 nt sequence from a lambda genomic clone, which was inserted between these two halves. The genomic organization of SHIP1 is shown in Wolf *et al.* (2000). The starting genomic clone contained a 4 kb region from the SacI site near the 3' end of the 7.6 kb

genomic clone in intron 6, extending through exon 8 and into intron 8. This SacI-SacI fragment was cloned into the SacI site of pBluescript SK (pBSK). The GFP gene from pEGFP-1 (Invitrogen/Clontech) was excised with NcoI (encompassing the ATG translation start site of GFP) and SspI. This was ligated into the NcoI (the putative s-SHIP translation start site in exon 7) and EcoRV sites of the pBSK-4kb clone. Next, the 5' half of the genomic promoter was added in the form of the SacI-SacI 7.6 kb genomic sub-clone. This was inserted into the one remaining SacI site at the 5' end of the intron 6-exon 7-GFP clone in pBSK. This left a gap of 0.9 kb between the two SacI sites in intron 6 (see Wolf *et al.*, 2000). This region was recovered as a larger BsiWI-EcoRI 2117 nt fragment, whose sequence demonstrated the insertion of 833 nucleotides between two SacI sites. Therefore, this BsiWI-EcoRI fragment was inserted into the same unique sites of the transgenic construct to produce the finished 11.5kb-GFP transgene in pBSK.

The 6.2kb-GFP transgene-construct was prepared from the 11.5kb-GFP transgene prior to the insertion of the 833 nt at the intron 6 SacI site. This 11.5kb(Δ833)-GFP construct was digested with FbaI and SwaI, removing 5.3 kb from the 5' end of intron 5. Re-ligation removed all but 19 intron 5 nt at the 5' end of the 11.5kb-GFP transgene. Both 11.5kb-GFP and 6.2kb-GFP transgenes, in pBSK, were cut from the plasmid with BssHII and Qiagen purified from an agarose gel for introduction into the mouse genome.

Production of transgenic mice

Founder transgenic mice were prepared in our Transgenic Mouse Facility by pronuclear injection of fertilized zygotes from (C57Bl/6 female X CBA/J male) F1 mice. Mice, positive for the transgene, were screened by PCR using DNA obtained from tails or toes of young animals. The location of the primer set for PCR is shown in FIG. 3: the upstream primer (a) is within intron 6 (Pro-up2, 5'-TACTCCTCAGCAAGAGTAGCTGG-3')(SEQ ID NO:XX), and the downstream primer (b) within the GFP gene (GFP-dn1, 5'-GCTGAACTTGTGGCCGTTTACGT-3')(SEQ ID NO:XX) produce a 632 nucleotide (nt) product. These primers were used for detection of both 6.2kb-GFP and 11.5kb-GFP transgenic mice. Positive chimeric mice were bred to C57Bl/6 mice and four founder lines (A, B, C and D) obtained for the 11.5kb-GFP mice. Later analyses demonstrated that founder line B was not positive for GFP

expression, even though the primer pair a and b gave a positive 632 nt product. Therefore, line B is not included in further analyses. The other lines were maintained by breeding transgene-positive animals with wild-type C57Bl/6 mice. For some experiments transgene-positive offspring were generated from positive intra-line breeding. Two founder animals were obtained for the 6.2kb-GFP transgene but one was lost.

The transgene copy number in each founder line (except 11.5kb-GFP, line B) was determined by semi-quantitative RT-PCR of transgene expression relative to endogenous *Gab2* expression. Primers for detecting genomic *gab2* are: E4F, 5'-CTTCTATAGCCTTCCCAAGCC-3' (SEQ ID NO:XX); E5R, 5'-CTCGTAGGTCTCACAGGAAG-3' (SEQ ID NO:XX).

Analysis of embryos

Preimplantation embryos were harvested at 2.5 and 3.5 dpc from uterine horns of pregnant females [see Nagy *et al.*, (2003) for details of these methods]. The morulae and blastocysts were washed in RPMI 1640 medium (Gibco) containing 10% fetal bovine serum, transferred to PBS (Ca^{2+} and Mg^{2+}), and GFP-expression or phase images photographed on a Nikon Eclipse TE200 inverted microscope coupled to a Roper Scientific 1kx1k pixel digital camera. Images were captured with MetaMorph software and prepared for publication with Photoshop (Adobe). High-resolution z-sections of GFP expression within embryos were made with a Leica TCS SP Confocal microscope.

Several blastocysts were plated onto gelatin-coated tissue-culture wells in DME 10% fetal bovine serum, and photographed three days later. During this period, blastocysts hatched from the zona pellucida, and attached to the culture plate. The attached mass of trophectoderm cells with the non-adherent ICM was photographed for GFP and phase with a Nikon Eclipse TE200 microscope.

RT-PCR analysis of s-SHIP expression in blastocysts

mRNA was isolated from wild-type 3.5 dpc blastocysts, FDC-P1 cells and the D3 ES cells using a Dynabeads mRNA DIRECT micro kit (Dynal). Reverse transcription used the Sensiscript kit from Qiagen, and the PCR cycling conditions were as follows: 94°C 1 min, [94°C 15 sec, 68°C 2 min] x 30 cycles, 68°C 5 min, and a 4°C hold. Each

reaction used the equivalent of 1.5 ng mRNA, based on the concentration before reverse transcription. Primers pairs were:

HPRT-up1, 5'-CCTGCTGGATTACATTAAAGCACTG-3' (SEQ ID NO:XX),
HPRT-down1 5'-GTCAAGGGCATATCCAACAACAAAC-3' (SEQ ID NO:XX);

OCT4-Up1 5'-GGCGTTCTCTTTGGAAAGGTGTTC-3' (SEQ ID NO:XX),
OCT4-Down1 5'-CTCGAACCACATCCTTCTCT-3' (SEQ ID NO:XX);

SHIP1/s-SHIP pair #3,
SHIP-E8FW, 5'-TTGCTGCACGAGGGCTCAGAATC-3' (SEQ ID NO:XX),
SSP883RV, 5'-TCCGATTCTCATGCTCTGGCTTG-3' (SEQ ID NO:XX);

SHIP1/s-SHIP pair #4,
SP2109FW, 5'-CAGCCCTGTCTTTGCCACGTTTG-3' (SEQ ID NO:XX),
SP2637RV, 5'-TCCACTGGATTCATCCCGCTCTG-3' (SEQ ID NO:XX);

SHIP1/s-SHIP pair #5,
newfw, 5'-CTTCCTCTTGCAACAGAGAACCC-3' (SEQ ID NO:XX),
newrv, 5'-ACTCAACGTCCACTTTGAGATGC-3' (SEQ ID NO:XX).

EXAMPLE 2:

Identification and Characterization of the *s-SHIP* Promoter

Potential *s-SHIP* promoter activity was first analyzed in cell lines grown in culture. Several cell lines were tested for s-SHIP vs. SHIP1 protein expression, based on the known and expected expression pattern of the s-SHIP protein (Lioubin *et al.*, 1994; Tu *et al.*, 2001). These results showed the expression of the ~104-kDa s-SHIP only in the ES cells, whereas the 145-kDa SHIP1 product was exclusively expressed in the maturing FD-Fms myeloid cells. Hot SDS-extraction of the ES cells did not change the size of the s-SHIP protein, suggesting that this 104-kDa product is not the result of proteolytic degradation during extraction (Horn *et al.*, 2001). SHIP proteins were not detectable in

NIH3T3 fibroblasts, the SNL cells serving as feeder for the ES cell growth, or the 293 human kidney cells. Therefore, NIH3T3 cells and D3 ES cells were selected as negative and positive cells, respectively, for analysis of the potential *s-SHIP* promoter activity.

A 7.6-kb genomic *ship1* region containing the intron-5 region was obtained for initial promoter analysis. The entire 7.6-kb region and sub-fragments thereof were cloned into a promoter-less GFP (enhanced green-fluorescent protein) expression vector (FIG. 1). Promoter activity of the intron-5 region was then assayed in the cells positive for s-SHIP expression (embryonic stem cells, clone D3) vs. cells negative for s-SHIP expression (NIH3T3 cells). The expression of GFP in each cell type, assayed by flow cytometry, was a measure of the promoter activity within each fragment of the 7.6 kb genomic DNA. The results indicated that, whereas, empty vectors alone lacked significant promoter activity in either cell type, vectors containing intron-5 segments exhibited substantial expression in the D3 ES cells but not in the NIH3T3 cells. Segments of intron 5, ranging from 0.96 kb to 7.6 kb were active for GFP expression in the ES cells; however, the shorter segments appeared most active. Two fragments of 1.9 kb and 0.96 kb, immediately upstream of exon 6, each exhibited equally high GFP expression. The shortest insert fragment contained part of exon 6, but only the 44 nucleotides upstream of exon 6, (Tu *et al.*, 2001), and was completely without promoter activity. These results strongly suggest that the intron-5 region of genomic *ship1* contains cell-specific promoter activity, and segments more distal to exon 6 may have negative regulatory activity.

Based on the ES/NIH3T3 cell-transfection experiments, two new constructs with an extended region downstream of the intron-5 genomic area were prepared for *in vivo* analysis of promoter activity in transgenic mice (FIG. 3A). Transgenic mice were produced for in vivo examination of the putative s-SHIP promoter/enhancer activity, and determining the overall expression pattern of the transgene, and presumably s-SHIP protein. The promoter in the longer of the new constructs (the 11.5kb-GFP transgene) contained the entire intron 5 from the above 7.6-kb genomic fragment, plus all of exon 6, intron 6, and the portion of exon 7 ending at the theoretical ATG start site (Kozak, 1987) for the s-SHIP protein translation. This start site was fused, in frame, to the ATG for the GFP protein. All of intron 6 and part of exon 7 were included in this construct because, 1)

the construct might then more closely resemble the endogenous promoter, 2) splicing may be important for efficient expression (Nott *et al.*, 2004), and 3) positive or negative regulatory elements for expression may also reside within this sequence. The second, shorter, transgenic promoter construct (the 6.2kb-GFP transgene) was similar, but contained only 0.96 kb of intron 5 sequence adjacent to exon 6, and also lacked 833 nucleotides between two SacI sites within intron 6. Thus, if either construct contained promoter activity *in vivo*, transcription would start within intron 5, while intron 6 would be spliced out and translation of GFP would begin at the first ATG within an appropriate Kozak site.

Transgenic (Tg) mice were then produced in the Hutchinson Center Transgenic Mouse facility and chimera animals screened for each transgene by PCR. Breeding each founder to wild-type C57Bl/6 mice yielded four lines containing the 11.5kb-GFP transgene, and one line with the 6.2kb-GFP transgene. Of the four founder Tg11.5 kb-GFP mice, one was negative for expression of the transgene (line B), while three were positive and each has exhibited the same expression patterns (lines A, C and D). Copy numbers of genomic transgenes, measured relative to the endogenous *gab2* gene are shown in FIG. 3B. Within the three GFP-expressing 11.5kb-GFP founder mice, empirical results indicate that line C exhibits the noticeably highest GFP expression levels. Line C mice also exhibit lower birth rates with *in utero* death at 8.5-9.5 days postcoitum (dpc) apparent. The single 6.2kb-GFP founder line harbors the most transgene copies, but no overt defects in the physical appearance of these mice, their birth rate or development have been observed.

Experiments were then conducted with the adult transgenic 11.5kb-GFP mice to examine transgene expression; however, it was difficult initially to find any GFP expressed in these mice by flow cytometry of blood and stem cell enriched bone marrow. After several negative attempts to find GFP expression, it was reasoned that because ES cell expression was readily detectable in the initial ES cell experiments, the best test for in vivo expression would be the inner cell mass (ICM) of the blastocyst, from which ES cells can be derived. Therefore, we looked for GFP expression in 3.5-dpc blastocysts derived from mating of Tg males x WT females. Blastocysts derived from one such cross produced 9 GFP-positive embryos indicating that the Tg was homozygous for the

transgene. A separate Tg male bred to a WT female produced both positive and negative blastocysts. GFP-positive morulae were also obtained from similar crosses; whereas, blastocysts or morulae from WT parents were negative for GFP.

Blastocysts are composed of 2-3 cell types depending on their developmental stage. The outer trophectoderm layer of cells surrounds the eccentric inner cell mass (ICM), destined to become the embryo proper, and later stage blastocysts also contain endodermal cells separating the ICM from the blastocoel cavity (Nagy *et al.*, 2003). To obtain a better idea of which cells of the blastocyst express the GFP transgene, transgenic 3.5-dpc blastocysts were allowed to adhere to a culture dish by three days growth in DME 10% FBS. Under these conditions, the zona pellucida is shed, and the outer trophectoderm cells of the blastocyst form an adherent layer while the ICM remains as an unspread mass, and each is distinguishable morphologically from the other. The results showed that the ICM portion of the blastocyst retained the GFP expression while the adherent trophectoderm cells were largely GFP-negative.

A more detailed picture of GFP expression throughout the intact early pre-implantation embryos was seen in confocal Z-sections of GFP within transgenic 2.5-dpc morulae and 3.5-dpc blastocysts. All cells of the 16 to 32-cell morula were GFP-positive. Transition of the morula to the early blastocyst is marked by the formation of the blastocoel cavity. A few cells of this early blastocyst structure began to shut-off GFP expression, and the extent of this GFP shut-off was more evident in the late blastocyst. Here, the outer trophectoderm cells had noticeably lower GFP expression, and the GFP-positive cells were confined to the ICM. Endodermal cells were not readily apparent. In these images, it is helpful to remember that the half-life of the GFP fluorescence is greater than 24 hr (Tech. Borchure, BD Bioscience ClonTech), and therefore cells, which have stopped expressing GFP, will retain some GFP protein and fluorescence for several days. Twenty-four hours separates the morula from the blastocyst stages; therefore, transgene shut-off early during this time would result in lower but not complete lack of GFP fluorescence late in this time span. The 11.5-kb transgene *s-SHIP* promoter contains the information for both cell-specific positive expression in morula and ICM of the blastocyst, but also cell-specific shut-off in trophectoderm cells.

Preimplantation embryos from the Tg6.2kb-GFP mice were analyzed next. The transgene in these mice contained only the proximal 0.96-kb region upstream of exon 6, which was necessary for GFP expression in the ES cells. It also lacked 833 nucleotides between two SacI sites of the intron-6 region. GFP expression in the 3.5-dpc blastocyst of the 6.2kb-GFP line was analyzed. Both qualitative and quantitative features of GFP expression in the Tg6.2kb-GFP blastocysts differed from those in the Tg11.5kb-GFP mice. First, GFP expression in the Tg6.2kb-GFP blastocysts was noticeably stronger (at least 5-fold) than that in the Tg11.5kb-GFP blastocysts, as measured by exposure times for obtaining equivalent GFP images in the Nikon digital microscope. Second, and more noticeable was the lack of GFP shut-off in the trophectoderm cells of the blastocyst. No clear demarcation in GFP expression was evident between ICM vs. trophectoderm as seen in the Tg11.5kb-GFP blastocysts.

Blastocysts from the Tg6.2kb-GFP mice were also allowed to adhere to culture plates and GFP expression was examined. Adherent blastocysts from Tg11.5kb-GFP mice were examined simultaneously. Adherent Tg6.2kb-GFP blastocyste expressed GFP in both ICM and trophectoderm cells in a, frequently, haphazard pattern. The Tg11.5kb-GFP adherent blastocysts expressed GFP only in the ICM as observed previously. A comparison of all embryos examined revealed that an increased GFP expression was apparent within the adherent Tg6.2kb-GFP blastocysts relative to the adherent Tg11.5kb-GFP blastocysts. These results were consistent with the promoter analyses performed in the ES cells (FIG. 1), and suggested that the lack of GFP shut-off by the 6.2kb-GFP transgene was due to negative regulatory information found in either one or both regions of the 11.5kb-GFP construct missing from the 6.2kb-GFP transgene.

The data from Tu *et al.* (2001) and that presented herein demonstrated exclusive s-SHIP (rather than SHIP1) expression in ES cells, yet, even though ES cells are derived from the ICM of the blastocyst and the intron 5 *s-SHIP* promoter functioned well in the ICM, it was still not certain whether the ICM actually expressed s-SHIP *in vivo*. Consequently, s-SHIP mRNA expression was then analyzed by RT-PCR, compared to that of the universally expressed HPRT, and the ES cell and ICM-specific Oct4 transcription factor. RNA from blastocysts, FDC-P1 myeloid progenitor cells, and D3 ES cells, was positive for HPRT as expected, and only the blastocysts and ES cells were

positive for Oct4. Initially, the s-SHIP-specific primers similar to those described by Tu *et al.* (2001) was used to test for s-SHIP expression; however, poor results were obtained. The forward primer in this set was moved 3'-ward into the region identical to SHIP1 but weak detection was still obtained. s-SHIP was therefore detected by "subtraction" using primers detecting both s-SHIP and SHIP1 products vs. primers detecting only the SHIP1 product. These primers clearly demonstrated the presence of full-length SHIP1 only in the FDC-P1 cells, and s-SHIP in both blastocysts and ES cells. The weak detectability of s-SHIP may be due to poor hybridization of the primers, degradation of the 5' s-SHIP mRNA ends, or possibly an additional shorter transcription product from the *ship1* gene.

Examination of the minimal 0.96-kb promoter proximal to exon 6 by MatInspector indicated at several transcription-factor binding site potentially active in ES cells and the blastocyst ICM. FIG. 4 shows the first 600 nucleotides of this region upstream of exon 6, with potential transcription factor binding sites and motifs for transcriptional regulation marked. A transcription initiator sequence (Butler and Kadonaga, 2002) straddles the 5' end of the 44 nt SSR, suggesting a transcriptional start site. Paired GATA, or Lmo2 binding sites are present, two overlapping p53 and Oct-binding sites, and a single extended FOX-factor binding region are prominent motifs. The Oct-binding motif is present in similar regions of both the murine and human s-SHIP promoter, suggesting such a factor could be important for ES and ICM expression. The POU factor Oct4 is expressed in ES cells and is part of an enhancer for ES cell-specific expression of target genes (Dailey *et al.*, 1994). Therefore the Oct site could be part of a similar ES cell enhancer region.

The transgene expression in preimplantation embryos raises a question about possible progenitor transgene expression in the oocytes or sperm of the adult, which then give rise to the fertilized embryo. The transcription factor, Oct4, is expressed in adult and embryonic germ cells, as well as the blastocyst ICM and in ES cells (Pesce *et al.*, 1998). The possibility that the 11.5kb-GFP transgene could also be germ cell specific is even more likely given the prominent Oct4 binding motif within the 0.96 kb minimal promoter upstream of exon 6 (see FIG. 4). Therefore, ovaries and testes from 7-8 week old adult Tg11.5kb-GFP mice were harvested and frozen sections stained with Alexa 594-labeled phalloidin for visualizing tissue structure through polymerized actin staining, and

endogenous GFP expression. The results of this experiment demonstrated that neither the developing sperm of the testis, nor the developing oocytes of the ovarian follicles expressed GFP. Only blood vessels of the testes and ovaries exhibited specific GFP expression. Therefore, unlike the Oct4 transcription factor, the 11.5kb-GFP transgene is not a maternally activated gene, must be transcriptionally activated sometime after the germ cells leave the ovary/testis, and before the 2.5-dpc-morula stage of development.

****

All of the compositions and/or methods and/or apparatus disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the compositions and/or methods and/or apparatus and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents that are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

REFERENCES

The following references are specifically incorporated herein by reference.

U.S. Patent 5,616,491

U.S. Patent 4,683,202

U.S. Patent 4,873,191

U.S. Patent 5,175,384

U.S. Patent 5,175,385

U.S. Patent 5,302,523

U.S. Patent 5,322,783

U.S. Patent 5,399,363

U.S. Patent 5,464,765

U.S. Patent 5,466,468

U.S. Patent 5,530,179

U.S. Patent 5,538,877

U.S. Patent 5,538,880

U.S. Patent 5,543,158

U.S. Patent 5,550,318

U.S. Patent 5,563,055

U.S. Patent 5,565,186

U.S. Patent 5,580,859

U.S. Patent 5,589,466

U.S. Patent 5,610,042

U.S. Patent 5,612,486

U.S. Patent 5,625,125

U.S. Patent 5,639,457

U.S. Patent 5,641,515

U.S. Patent 5,656,610

U.S. Patent 5,702,932

U.S. Patent 5,736,524

U.S. Patent 5,780,448
U.S. Patent 5,789,215
U.S. Patent 5,846,225
U.S. Patent 5,846,233
U.S. Patent 5,925,565
U.S. Patent 5,928,906
U.S. Patent 5,935,819
U.S. Patent 5,945,100
U.S. Patent 5,981,274
U.S. Patent 5,994,624
U.S. Patent 4,237,224
U.S. Patent 5,384,253

Abbott *et al., J. Nucl. Cardiol.,* 10(4):403-412, 2003.
Abremski and Hess, *J. Mol. Biol.*, 259:1509-1514, 1984.
Abremski *et al., Cell,* 32:1301-1311, 1983.
Almendro *et al., J. Immunol.*, 157(12):5411-5421, 1996.
Amin *et al., J. Molec. Biol.*, 214:55-72, 1990.
Anderson *et al., J. of Immunotherapy,* 12:19-31, 1992.
Angel *et al., Mol. Cell. Biol.*, 7:2256, 1987.
Angel *et al., Cell,* 49:729, 1987b.
Angel *et al., Mol. Cell. Biol.*, 7:2256, 1987a.
Armentano *et al., Proc. Natl. Acad. Sci. USA,* 87(16):6141-6145, 1990.
Atchison and Perry, *Cell,* 46:253, 1986.
Atchison and Perry, *Cell,* 48:121, 1987.
Ausubel *et al., In: Current Protocols in Molecular Biology,* John, Wiley & Sons, Inc, NY, 1996.
Ausubel *et al., In: Current Protocols in Molecular Biology,* John, Wiley & Sons, Inc, NY, 1994.
Ausubel *et al., In: Current Protocols in Molecular Biology,* John, Wiley & Sons, Inc, NY, 1992.

Baichwal and Sugden, *In: Gene Transfer,* Kucherlapati (Ed.), NY, Plenum Press, 117-148, 1986.

Banerji *et al.*, *Cell*, 27(2 Pt 1):299-308, 1981.

Banerji *et al.*, *Cell*, 33(3):729-740, 1983.

Bellus, *J. Macromol. Sci. Pure Appl. Chem.*, A31(1): 1355-1376, 1994.

Berkhout *et al.*, *Cell*, 59:273-282, 1989.

Blanar *et al.*, *EMBO J.*, 8:1139, 1989.

Bodine and Ley, *EMBO J.*, 6:2997, 1987.

Boshart *et al.*, *Cell*, 41:521, 1985.

Bosze *et al.*, *EMBO J.*, 5(7):1615-1623, 1986.

Braddock *et al.*, *Cell*, 58:269, 1989.

Brinster *et al.*, *Proc. Natl. Acad. Sci. USA*, 82(13):4438-4442, 1985.

Bulla and Siddiqui, *J. Virol.*, 62:1437, 1986.

Butler and Kadonaga, *Genes Dev.*, 16:2583-2592, 2002.

Campbell and Villarreal, *Mol. Cell. Biol.*, 8:1993, 1988.

Campere and Tilghman, *Genes and Dev.*, 3:537, 1989.

Campo *et al.*, *Nature*, 303:77, 1983.

Capaldi *et al.*, *Biochem. Biophys. Res. Comm.*, 74(2):425-433, 1977.

Carbonelli *et al.*, *FEMS Microbiol. Lett.*, 177(1):75-82, 1999.

Celander and Haseltine, *J. Virology*, 61:269, 1987.

Celander *et al.*, *J. Virology,* 62:1314, 1988.

Chandler *et al.*, Cell, 33:489, 1983.

Chandler *et al.*, *Proc. Natl. Acad. Sci. USA*, 94(8):3596-601, 1997.

Chang *et al.*, *Mol. Cell. Biol.*, 9:2153, 1989.

Chatterjee *et al.*, *Proc. Natl. Acad. Sci. USA*, 86:9114, 1989.

Chen and Okayama, *Mol. Cell Biol.*, 7(8):2745-2752, 1987.

Choi *et al.*, *Cell*, 53:519, 1988.

Chowdhury *et al.*, *Science*, 254(5039):1802-1805, 1991.

Cocea, *Biotechniques*, 23(5):814-816, 1997.

Cohen *et al.*, *J. Cell. Physiol.*, 5:75, 1987.

Costa *et al.*, *Mol. Cell. Biol.*, 8:81, 1988.

Coupar *et al.*, *Gene*, 68:1-10, 1988.

Cournoyer *et al.*, *Curr. Opin. Biotechnol.*, 1(2):196-208, 1990.

Craig, *Ann. Rev. Genet.*, 22:77-105, 1988.

Cregg and Madden, *Mol. Gen. Genet.*, 219:320-323, 1989.

Cripe *et al.*, *EMBO J.*, 6:3745, 1987.

Cristiano *et al.*, *Proc. Natl. Acad. Sci. USA*, 90(24):11548-11552, 1993.

Culotta and Hamer, *Mol. Cell. Biol.*, 9:1376, 1989.

Dai *et al.*, *J. Biol. Chem.*, 267(27):19565-19571, 1992.

Dailey *et al.*, *Mol. Cell. Biol.*, 14:7758-7769, 1994.

Daley *et al.*, *Exp. Hematol.*, 21(6):734-737, 1993.

Dandolo *et al.*, *J. Virology*, 47:55-64, 1983.

de Villartay, *Nature*, 335:170-174, 1988.

De Villiers *et al.*, *Nature*, 312(5991):242-246, 1984.

Deschamps *et al.*, *Science*, 230:1174-1177, 1985.

Echols, *J. Biol. Chem.*, 265:14697-14700, 1990.

Edbrooke *et al.*, *Mol. Cell. Biol.*, 9:1908, 1989.

Edlund *et al.*, *Science*, 230:912-916, 1985.

Fechheimer, *et al.*, *Proc Natl. Acad. Sci. USA*, 84:8463-8467, 1987.

Feng and Holland, *Nature*, 334:6178, 1988.

Ferry *et al.*, *Proc. Natl. Acad. Sci. USA*, 88(19):8377-8381, 1991.

Firak and Subramanian, *Mol. Cell. Biol.*, 6:3667, 1986.

Flanagan and Fennewald, *J. Molec. Biol.*, 206:295-304, 1989.

Foecking and Hofstetter, *Gene*, 45(1):101-105, 1986.

Fraley *et al.*, *Proc. Natl. Acad. Sci. USA*, 76:3348-3352, 1979.

Friedmann, *Science*, 244:1275-1281, 1989.

Friedmann, *Science*, 244:1275-1281, 1989.

Fujita *et al.*, Cell, 49:357, 1987.

Fung-Leung *et al.*, *Cell*, 65(3):443-449, 1991b.

Fung-Leung *et al.*, *J Exp Med.*, 174(6):1425-1429, 1991a.

Gilles *et al.*, *Cell*, 33:717, 1983.

Glasgow *et al.*, *J. Biol. Chem.*, 264:10072-10082, 1989.

Gloss *et al.*, *EMBO J.*, 6:3735, 1987.

Godbout *et al.*, *Mol. Cell. Biol.*, 8:1169, 1988.

Golic and Lindquist, *Cell*, 59:499-509, 1989.

Goodbourn and Maniatis, *Proc. Natl. Acad. Sci. USA*, 85:1447, 1988.

Goodbourn *et al.*, *Cell*, 45:601, 1986.

Gopal, *Mol. Cell Biol.*, 5:1188-1190, 1985.

Graham and Van Der Eb, *Virology*, 52:456-467, 1973.

Greene *et al.*, *Immunology Today*, 10:272, 1989

Grosschedl and Baltimore, *Cell*, 41:885, 1985.

Gupta *et al.*, *J. Biol. Chem.*, 274:7489-7494, 1999.

Haffter and Bickle, *EMBO J.*, 7:3991-3996, 1988.

Harland and Weintraub, *J. Cell Biol.*, 101(3):1094-1099, 1985.

Haslinger and Karin, *Proc. Natl. Acad. Sci. USA*, 82:8572, 1985.

Hauber and Cullen, *J. Virology*, 62:673, 1988.

Hen *et al.*, *Nature*, 321:249, 1986.

Hensel *et al.*, *Lymphokine Res.*, 8:347, 1989.

Hermonat and Muzycska, *Proc. Natl. Acad. Sci. USA*, 81:6466-6470, 1984.

Herr and Clarke, *Cell*, 45:461, 1986.

Herz and Gerard, *Proc. Natl. Acad. Sci. USA*, 90:2812-2816, 1993.

Hirai, *Hum Cell.*, 15(4):190-198, 2002.

Hirochika *et al.*, *J. Virol.*, 61:2599, 1987.

Hirsch *et al.*, *Mol. Cell. Biol.*, 10:1959, 1990.

Hogan *et al.*, In: Manipulating the Mouse Embryo: A Laboratory Manual, 2nd ed.,Cold Spring Harbor Laboratory Press, 1994.

Holbrook *et al.*, *Virology*, 157:211, 1987.

Horlick and Benfield, *Mol. Cell. Biol.*, 9:2396, 1989.

Horn *et al.*, *Leukemia*, 15:112-120, 2001.

Horwich *et al. J. Virol.*, 64:642-650, 1990.

Huang *et al.*, *Cell*, 27:245, 1981.

Hubner *et al.*, *J. Molec. Biol.*, 205:493-500, 1989.

Hug *et al.*, *Mol. Cell. Biol.*, 8:3065, 1988.

Hunger-Bertling *et al., Mol. Cell. Biochem.*, 92:107-116, 1990.

Hwang *et al., Mol. Cell. Biol.*, 10:585, 1990.

Hwu *et al., J. Immunol.*, 150(9):4104-4115, 1993.

Imagawa *et al., Cell*, 51:251, 1987.

Imbra and Karin, *Nature*, 323:555, 1986.

Imler *et al., Mol. Cell. Biol.*, 7:2558, 1987.

Imperiale and Nevins, *Mol. Cell. Biol.*, 4:875, 1984.

Inouye and Inouye, *Nucleic Acids Res.*, 13:3101-3109, 1985.

Jakobovits *et al., Mol. Cell. Biol.*, 8:2555, 1988.

Jameel and Siddiqui, *Mol. Cell. Biol.*, 6:710, 1986.

Jaynes *et al., Mol. Cell. Biol.*, 8:62, 1988.

Johnson *et al., Mol. Cell. Biol.*, 9:3393, 1989.

Kadesch and Berg, *Mol. Cell. Biol.*, 6:2593, 1986.

Kaeppler *et al., Plant Cell Reports*, 9:415-418, 1990.

Kaneda *et al., Science*, 243:375-378, 1989.

Karin *et al., Mol. Cell. Biol.*, 7:606, 1987.

Karin *et al., Mol. Cell. Biol.*, 7:606, 1987.

Katinka *et al., Cell*, 20:393, 1980.

Kato *et al, J. Biol. Chem.*, 266:3361-3364, 1991.

Kavanaugh *et al., Current Biol.*, 6:438-445, 1996.

Kawamoto *et al., Mol. Cell. Biol.*, 8:267, 1988.

Kay *et al., Hum. Gene Ther.*, 3(6):641-647, 1992.

Kiledjian *et al., Mol. Cell. Biol.*, 8:145, 1988.

Klamut *et al., Mol. Cell. Biol.*, 10:193, 1990.

Koch *et al., Mol. Cell. Biol.*, 9:303, 1989.

Kondo *et al., Annu. Rev. Immunol.*, 21:759-806., 2003.

Kozak, *Nucleic Acids Res.*, 15:8125-8148, 1987.

Kraus *et al. FEBS Lett.*, 428(3):165-170, 1998.

Kriegler and Botchan, *In: Eukaryotic Viral Vectors*, Gluzman (Ed.), Cold Spring Harbor: Cold Spring Harbor Laboratory, NY, 1982.

Kriegler and Botchan, *Mol. Cell. Biol.*, 3:325, 1983.

Kriegler *et al.*, *Cell*, 38:483, 1984.

Kriegler *et al.*, *Cell*, 53:45, 1988.

Kuhl *et al.*, *Cell*, 50:1057, 1987.

Kunkel *et al.*, *Methods Enzymol.*, 154:367-382, 1987.

Kunz *et al.*, *Nucl. Acids Res.*, 17:1121, 1989.

Lareyre *et al.*, *J. Biol. Chem.*, 274(12):8282-8290, 1999.

Larsen *et al.*, *Proc Natl. Acad. Sci. USA.*, 83:8283, 1986.

Laspia *et al.*, *Cell*, 59:283, 1989.

Latimer *et al.*, *Mol. Cell. Biol.*, 10:760, 1990.

Lee *et al.*, *Biochem. Biophys. Res. Commun.*, 240(2):309-313, 1997.

Lee *et al.*, *Nature*, 294:228, 1981.

Lee *et al.*, *Nucleic Acids Res.*, 12:4191-206, 1984.

Levenson *et al.*, *Hum. Gene Ther.*, 9(8):1233-1236, 1998.

Levinson *et al.*, *Nature*, 295:79, 1982.

Lin *et al.*, *Mol. Cell. Biol.*, 10:850, 1990.

Lioubin *et al.*, *Genes Devel.*, 10:1084-1095, 1996.

Lioubin *et al.*, *Mol. Cell Biol.*, 14(9):5682-5691, 1994.

Liu *et al.*, *Blood*, 91:2753-2759, 1998.

Liu *et al.*, *Mol. Cell. Biol.*, 21:3047-3056, 2001.

Lucas and Rohrschneider, *Blood*, 93:1922-1933, 1999.

Luria *et al.*, *EMBO J.*, 6:3307, 1987.

Lusky and Botchan, *Proc. Natl. Acad. Sci. USA*, 83:3609, 1986.

Lusky *et al.*, *Mol. Cell. Biol.*, 3:1108, 1983.

Macejak and Sarnow, *Nature*, 353:90-94, 1991.

Majors and Varmus, *Proc. Natl. Acad. Sci. USA*, 80:5866, 1983.

Malynn *et al.*, *Cell*, 54:453-460, 1988.

Matsuzaki *et al.*, *J. Bacteriol.*, 172:610-618, 1990.

McNeall *et al.*, *Gene*, 76:81, 1989.

Mercier *et al.*, *J. Bacteriol.*, 172:3745-757, 1990.

Miksicek *et al.*, *Cell*, 46:203, 1986.

Miller *et al.*, *Am. J. Clin. Oncol.*, 15(3):216-221, 1992.

Mordacq and Linzer, *Genes and Dev.*, 3:760, 1989.

Moreau *et al.*, *Nucl. Acids Res.*, 9:6047, 1981.

Muesing *et al.*, *Cell*, 48:691, 1987.

Nagy *et al.*, In: *Manipulating the Mouse Embryo,* A Laboratory Manual, Third Ed. Cold Spring Harbor Press, 2003.

Nakano *et al., Science*, 265:1098-1101, 1994.

Nakano, *Trends Immunol.,* 24(11):589-594, 2003.

Ng *et al.*, *Nuc. Acids Res.*, 17:601, 1989.

Nicolas and Rubenstein, In: *Vectors: A survey of molecular cloning vectors and their uses*, Rodriguez and Denhardt (Eds.), Stoneham: Butterworth, 494-513, 1988.

Nicolau and Sene, *Biochim. Biophys. Acta*, 721:185-190, 1982.

Nicolau *et al.*, *Methods Enzymol.*, 149:157-176, 1987.

Nomoto *et al.*, *Gene*, 236(2):259-271, 1999.

Nott *et al., Genes Dev.*, 18:210-222, 2004.

Ondek *et al.*, *EMBO J.*, 6:1017, 1987.

Ornitz *et al.*, *Mol. Cell. Biol.*, 7:3466, 1987.

Palmiter *et al.*, *Nature*, 300:611, 1982.

Parsons *et al.*, *J. Biol. Chem.*, 265:4527-4533, 1990.

PCT Appln. WO 94/09699

PCT Appln. WO95/06128

Pech *et al.*, *Mol. Cell. Biol.*, 9:396, 1989.

Pelletier and Sonenberg, *Nature*, 334(6180):320-325, 1988.

Perez-Stable and Constantini, *Mol. Cell. Biol.*, 10:1116, 1990.

Pesce *et al., Bioessays*, 20:722-732, 1998.

Pfeffer *et al., Cell,* 73(3):457-467, 1993.

Picard and Schaffner, *Nature*, 307:83, 1984.

Pinkert *et al.*, *Genes and Dev.,* 1:268, 1987.

Ponta *et al.*, *Proc. Natl. Acad. Sci. USA*, 82:1020, 1985.

Porton *et al.*, *Mol. Cell. Biol.*, 10:1076, 1990.

Potrykus *et al., Mol. Gen. Genet.*, 199:183-188, 1985.

Poyart-Salmeron *et al.*, *EMBO J.*, 8:2425-2433, 1989.

Quantin *et al., Proc. Natl. Acad. Sci. USA,* 89(7):2581-2584, 1992.

Queen and Baltimore, *Cell*, 35:741, 1983.

Quinn *et al.*, *Mol. Cell. Biol.*, 9:4713, 1989.

Redondo *et al.*, *Science*, 247:1225, 1990.

Reisman and Rotter, *Mol. Cell. Biol.*, 9:3571, 1989.

Remington's Pharmaceutical Sciences, 15th ed., pages 1035-1038 and 1570-1580, Mack Publishing Company, Easton, PA, 1980.

Resendez Jr. *et al.*, *Mol. Cell. Biol.*, 8:4579, 1988.

Ridgeway, *In: Vectors: A survey of molecular cloning vectors and their uses*, Rodriguez and Denhardt (Eds.), Stoneham:Butterworth, 467-492, 1988.

Ripe *et al.*, *Mol. Cell. Biol.*, 9:2224, 1989.

Rippe *et al.*, *Mol. Cell Biol.*, 10:689-695, 1990.

Rittling *et al.*, *Nuc. Acids Res.*, 17:1619, 1989.

Rohrschneider *et al., Genes Devel.*, 14:505-520, 2000.

Rohrschneider, In: *Handbook of Cell Signaling*, Bradshaw andDennis (Eds.), 148(2):147-151, Elsevier Sciences (USA), 2003.

Rosen *et al.*, *Cell,* 41:813, 1988.

Rosenfeld *et al.*, *Cell*, 68:143-155, 1992.

Sakai *et al.*, *Genes and Dev.,* 2:1144, 1988.

Sambrook *et al.*, *In: Molecular cloning,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

Satake *et al.*, *J. Virology,* 62:970, 1988.

Sato *et al.*, *J. Bacteriol.*, 172:1092-1098, 1990.

Sattler *et al., J. Biol. Chem.*, 276:2451-2458, 2001.

Schaffner *et al.*, *J. Mol. Biol.*, 201:81, 1988.

Schwartz and Sadowski, *J. Molec. Biol.*, 205:647-658, 1989.

Searle *et al.*, *Mol. Cell. Biol.*, 5:1480, 1985.

Sharp and Marciniak, *Cell*, 59:229, 1989.

Shaul and Ben-Levy, *EMBO J.*, 6:1913, 1987.

Sherman *et al.*, *Mol. Cell. Biol.*, 9:50, 1989.

Sleigh and Lockett, *J. EMBO*, 4:3831, 1985.

Sly *et al., Exp. Hematol.*, 31:1170-1181, 2003.

Spalholz *et al., Cell,* 42:183, 1985.

Spandau and Lee, *J. Virology*, 62:427, 1988.

Spandidos and Wilkie, *EMBO J.*, 2:1193, 1983.

Stephens and Hentschel, *Biochem. J.*, 248:1, 1987.

Sternberg *et al.*. Cold Spring Harbor Symp. Quant. Biol.. 45:297-309, 1981.

Stuart *et al., Nature*, 317:828, 1985.

Sullivan and Peterlin, *Mol. Cell. Biol.*, 7:3315, 1987.

Swartzendruber and Lehman, *J. Cell. Physiology*, 85:179, 1975.

Takebe *et al., Mol. Cell. Biol.*, 8:466, 1988.

Tamir *et al., Immunity*, 12:347-358, 2000.

Tavernier *et al., Nature,* 301:634, 1983.

Taylor and Kingston, *Mol. Cell. Biol.*, 10:165, 1990a.

Taylor and Kingston, *Mol. Cell. Biol.*, 10:176, 1990b.

Taylor *et al., J. Biol. Chem.*, 264:15160, 1989.

Temin, *In*: *Gene Transfer*, Kucherlapati (Ed.), NY, Plenum Press, 149-188, 1986.

Thiesen *et al., J. Virology,* 62:614, 1988.

Tomic *et al., Nucl. Acids Res.,* 12:1656, 1990.

Treisman, *Cell*, 42:889, 1985.

Tronche *et al., Mol. Biol. Med.*, 7:173, 1990.

Trudel and Constantini, *Genes and Dev.* 6:954, 1987.

Tsumaki *et al., J. Biol. Chem.,* 273(36):22861-22864, 1998.

Tu *et al., Blood*, 98:2028-2038, 2001.

Tyndell *et al., Nuc. Acids. Res.*, 9:6231, 1981.

Upender *et al., Biotechniques*, 18:29-31, 1995.

Van Beusechem *et al., Proc. Natl. Acad. Sci. USA,* 89(16):7640-7644, 1992.

Vannice and Levinson, *J. Virology*, 62:1305, 1988.

Vasseur *et al., Proc Natl. Acad. Sci. U.S.A.*, 77:1068, 1980.

Wang and Calame, *Cell*, 47:241, 1986.

Watson and Ramstad, In: *Corn: Chemistry and Technology*, 1987.

Weber *et al., Cell*, 36:983, 1984.

Weinberger *et al. Mol. Cell. Biol.*, 8:988, 1984.

Weisberg *et al.*, In: *Lambda II*, Hendrix *et al.* (Eds.), Cold Spring Harbor Press, Cold Spring Harbor, NY, 211-250, 1983.

Winoto and Baltimore, *Cell* 59:649, 1989.

Wolf *et al., Genomics,* 69:104-112, 2002.

Wolff *et al., Science,* 246:1464-1468, 1990.

Wong *et al., Gene*, 10:87-94, 1980.

Wu *et al., Biochem. Biophys. Res. Commun.,* 233(1):221-6, 1997.

Yutzey *et al. Mol. Cell. Biol.*, 9:1397, 1989.

Zhao-Emonet *et al., Biochim. Biophys. Acta*, 1442(2-3):109-119, 1998.

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising an *s-ship* promoter capable of promoting transcription.

2. The isolated polynucleotide of claim 1, wherein the promoter comprises at least 20 contiguous nucleotides from SEQ ID NO:1.

3. The isolated polynucleotide of claim 2, wherein the promoter comprises at least 50 nucleotides from SEQ ID NO:1.

4. The isolated polynucleotide of claim 3, wherein the promoter comprises at least 100 nucleotides from SEQ ID NO:1.

5. The isolated polynucleotide of claim 4, wherein the promoter comprises at least 500 nucleotides from SEQ ID NO:1.

6. The isolated polynucleotide of claim 5, wherein the promoter comprises at least 1000 nucleotides from SEQ ID NO:1.

7. The isolated polynucleotide of claim 6, wherein the promoter comprises at least 5000 nucleotides from SEQ ID NO:1.

8. The isolated polynucleotide of claim 7, wherein the promoter comprises about 6.3 kilobases from SEQ ID NO:1.

9. The isolated polynucleotide of claim 8, wherein the promoter comprises about 7.6 kilobases from SEQ ID NO:1.

10. The isolated polynucleotide of claim 2, comprising SEQ ID NO:1.

11. The isolated polynucleotide of claim 1, wherein the promoter is capable of promoting tissue-specific transcription.

12. The isolated polynucleotide of claim 1, wherein the promoter region is operably connected to a heterologous nucleic acid.

13. A nucleic acid comprising a promoter operably attached to a nucleic acid sequence from an *s-ship* gene or a portion thereof and a marker sequence, wherein the *s-ship* gene is disrupted by the marker sequence.

14. The nucleic acid of claim 13, wherein the promoter is an *s-ship* promoter.

15. The nucleic acid of claim 13, wherein the promoter is constitutive.

16. The nucleic acid of claim 13, wherein the promoter is inducible or conditional.

17. An expression cassette comprising an *s-ship* promoter operably connected to a nucleic acid segment.

18. The expression cassette of claim 17, wherein the nucleic acid segment is heterologous.

19. The expression cassette of claim 18, wherein the nucleic acid segment is a reporter gene.

20. The expression cassette of claim 19, wherein the reporter gene encodes a gene product that is colorimetric, enzymatic, or fluorescent.

21. The expression cassette of claim 18, wherein the nucleic acid segment encodes a therapeutic or diagnostic gene product.

22. The expression cassette of claim 21, wherein the therapeutic or diagnostic gene product is a polypeptide.

23. The expression cassette of claim 21, wherein the therapeutic or diagnostic gene product is an RNA molecule.

24. The expression cassette of claim 23, wherein the RNA molecule is an siRNA or miRNA molecule.

25. The expression cassette of claim 21, wherein the nucleic acid segment encodes a therapeutic gene product.

26. The expression cassette of claim 25, wherein the therapeutic gene product is selected from the group consisting of a tumor suppressor, a cytokine, a cytokine receptor, a differentiation-inducer, growth factor, and a growth factor receptor.

27. A vector comprising an *s-ship* promoter.

28. The vector of claim 1, wherein the *s-ship* promoter is operably attached to a nucleic acid segment.

29. The vector of claim 28, wherein the nucleic acid segment is all or part of an *s-ship1* coding sequence.

30. The vector of claim 28, wherein the nucleic acid segment is heterologous.

31. The vector of claim 27, wherein the vector is a plasmid, YAC, BAC, or virus.

32. The vector of claim 27, comprised in a pharmaceutically acceptable formulation.

33. A host cell comprising an *s-ship* promoter operably attached to a heterologous nucleic acid segment.

34. The host cell of claim 33, wherein the host cell is eukaryotic.

35. The host cell of claim 34, wherein the host cell is an embryonic cell.

36. The host cell of claim 35, wherein the embryonic cell is a blastocyst cell.

37. The host cell of claim 34, wherein the host cell is a hematopoietic cell.

38. The host cell of claim 34, wherein the host cell is a stem or progenitor cell.

39. The host cell of claim 38, wherein the stem or progenitor cell is from tissue selected from a group consisting of skin, a hair follicle, cornea, embryo, gonads, mammary gland, pancreas, and smooth muscle.

40. A recombinant host cell in which one or both *s-ship* genes is disrupted by marker sequence.

41. A transgenic animal comprising an *s-ship* promoter region operably attached to a heterologous nucleic acid segment.

42. The transgenic animal of claim 41, which is a mammal.

43. A mammal having cells comprising an *s-ship* transgenic sequence.

44. The mammal of claim 43, wherein the *s-ship* transgenic sequence comprises a *s-ship1* coding sequence flanked by loxP sequences.

45. The mammal of claim 44, further comprising a heterologous nucleic acid sequence encoding a Cre recombinase.

46. The mammal of claim 45, wherein the nucleic acid sequence encoding the Cre recombinase is under the control of an inducible or conditional promoter.

47. A method for expressing a recombinant nucleic acid in a stem or progenitor cell comprising:

a) transfecting the cell with an expression cassette comprising an *s-ship* promoter operably attached to the recombinant nucleic acid, wherein the nucleic acid is transcribed.

48. A method of screening for a candidate substance that regulates activity of the *s-ship1* promoter comprising a step selected from the group consisting of:

(a) contacting a nucleic acid comprising an *s-ship* promoter with an *s-ship* promoter binding protein and the candidate substance under conditions that allow binding between the protein and the promoter and determining

whether the candidate compound modulates the binding between the protein and the promoter; and

(b) contacting the candidate substance with a cell comprising the *s-ship* promoter operably attached to a reporter gene coding for an expression product and assaying for expression of the reporter gene expression product.

49. A method for identifying stem cells in a population of cells comprising:

(a) administering to cells in the population a nucleic acid comprising an *s-ship* promoter operably attached to a reporter gene.

50. The method of claim 49, wherein the cells are in an organ.

51. The method of claim 49, wherein the cell are in an animal.

52. The method of claim 49, further comprising sorting cells based on expression of the reporter gene.

53. A method for screening for a modulator of cell function comprising:

a) transfecting a stem or hematopoietic cell with an expression cassette comprising an *s-ship* promoter operably attached to a nucleic acid encoding a candidate modulator; and,

b) assaying the cell for a cell function, wherein a difference in cell function in the cell as compared to a cell in the absence of the candidate modulator is indicative of a modulator.

54. The method of claim 53, wherein the modulator is a candidate therapeutic agent for the treatment of a blood-related disease or condition.

55. A method of treating a patient with a blood-related disease or condition comprising:

a) transfecting a cell with an expression cassette comprising an *s-ship* promoter region operably attached to a therapeutic nucleic acid; and,

b) administering the cell to the patient.

56. The method of claim 55, wherein the cell is a bone marrow cell.

57. The method of claim 55, wherein the cell is autologous.

58. The method of claim 55, wherein the cell is allogeneic.

59. The method of claim 55, wherein the blood-related disease or condition is a blood-related cancer.

60. The method of claim 59, wherein the blood-related cancer is leukemia, lymphoma, or myeloma.

61. The method of claim 55, wherein the blood-related condition is anemia.

62. The method of claim 55, wherein the blood-related condition can be treated with stem cell replacement therapy.

ABSTRACT

The present invention concerns *s-ship* promoter compositions and methods using the promoter. It includes polynucleotides, vectors, host cells, and transgenic animal including an *s-ship* promoter controlling the expression of a heterologous nucleic acid. Methods of the invention concern methods of expressing a heterologous nucleic acid is a tissue-specific, developmental-specific, or temporally controlled manner. Other methods includes screening methods and therapeutic methods.

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

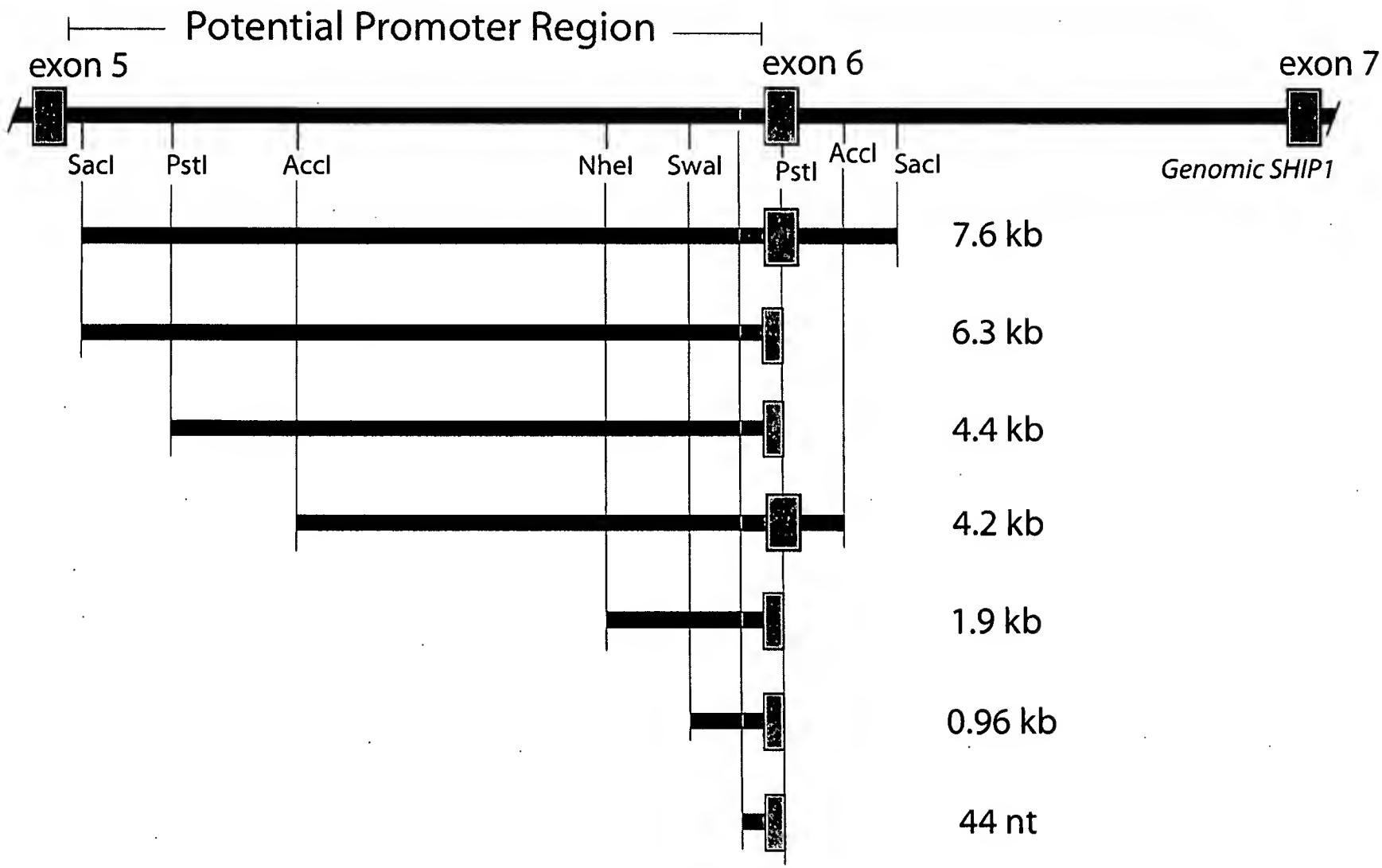
Potential Promoter Region
exon 5
exon 6
exon 7
SacI
PstI
AccI
NheI
SwaI
PstI
AccI
SacI
Genomic SHIP1
7.6 kb
6.3 kb
4.4 kb
4.2 kb
1.9 kb
0.96 kb
44 nt

FIG. 1

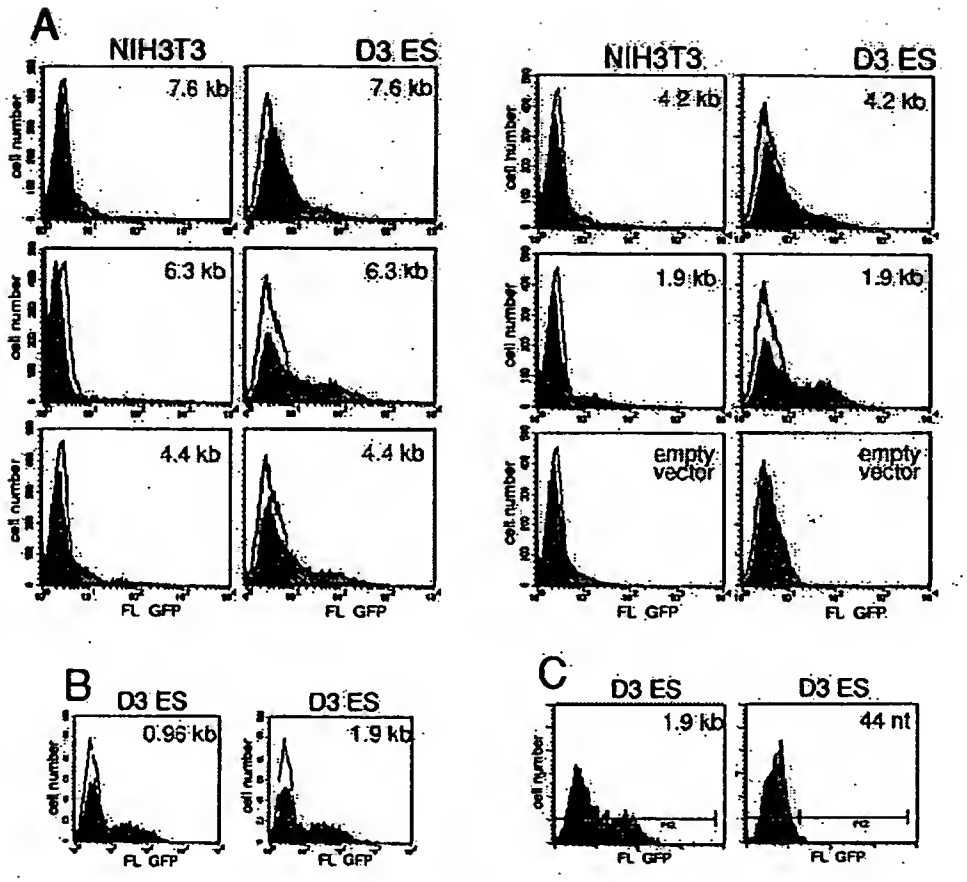
A
NIH3T3
D3 ES
7.6 kb
6.3 kb
4.4 kb
4.2 kb
1.9 kb
empty vector
cell number
FL GFP
B
D3 ES
0.96 kb
1.9 kb
C
D3 ES
1.9 kb
44 nt
M2

FIG. 2

FIG. 3A

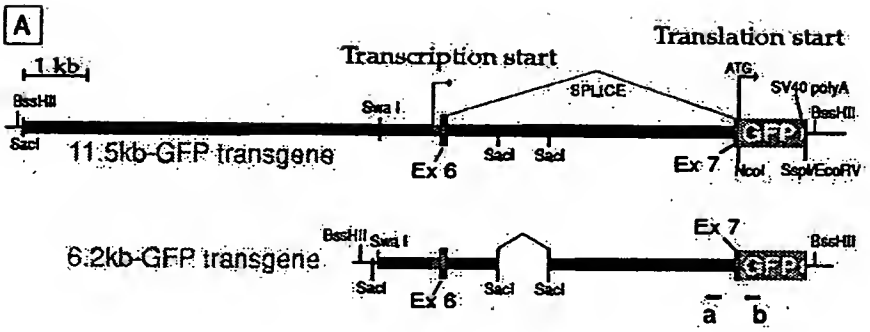

FIG. 3B

C TRANSENE COPY NUMBERS

transgene	Line	# copies
11.5 kb-GFP	A	5
"	C	13
"	D	11
6.2 kb-GFP	Z	17


-600 AGAAGTGATCTGGGCTGGACAGACTAGCTGAACTGGCCAGCTCTGGGTTC
-550 ATCAAGAAACCCTACCTCCATAACATAAAGTGTGATGGAGAAAGGCACCT
Areb6
-500 AATGTCAACCTCAAACCCCTACCTGCATGTGCACACACATACATCCACAC
-450 CACACACACACACACACACACACACACACCACACACACACACACACAA
Nkx3.1
Fox factor binding sites
-400 ATAAATAAGTAAATAAATAAAATATTTAGCTCTCCAGACCAAATCTTGGT
Oct4
Pax8
-350 GAAACCCATGCATTTGCATTTGTGTGTGTCCTACAAACACTGAAGGTTAA
Cdx2
-300 GAAGCATGCTCCTTAGTAATTTTATAGCAGTTTGCGTTTCCAGATTGAAA
Gata / Lmo2
-250 ACAGATTCTATAGGCTACACAGTGCTAAATGGATTATGCTCAGATACAGA
Smad3/4
-200 TTGAAAAGGATACAGATTGAAAAGGGTCGGGGTCTGGGCCAGGATGACGG
p53
Stat1/5
-150 GCCAACTATCTTTGCCCGGGCTTGTCCTTCAGGGAAGGGTTACAGGATTC
Gata / Lmo2
-100 ACCACTGGGGTGTGGCCTATCTGCTGTTAGGACCTGAATTGCCTGGAGTG
Initiator
stem-SHIP region
-50 TTTCTAGTTCCCACTAGTTGTTGAACTTTACCTTGAACCTCTGCTCCCAG

FIG. 4

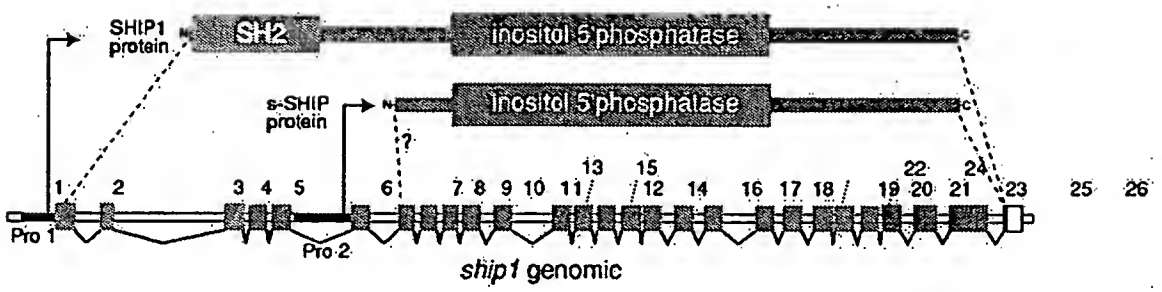
SHIP1 protein
SH2
inositol 5'phosphatase
s-SHIP protein
inositol 5'phosphatase
?
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Pro 1
Pro 2
ship1 genomic

FIG. 5

SEQUENCE LISTING

<110> ROHRSCHNEIDER, LARRY R.

<120> METHODS AND COMPOSITIONS INVOLVING S-SHIP PROMOTER REGIONS

<130> FHCC:016USP1

<140> UNKNOWN
<141> 2004-03-18

<160> 15

<170> PatentIn Ver. 2.1

<210> 1
<211> 100140
<212> DNA
<213> Mus musculus

<220>
<221> modified_base
<222> (27350)..(78168)
<223> N = A, C, G OR T/U

<400> 1
ggcaatttct gagaggcaac aggcggcagg tctcagccta gagagggccc tgaactactt 60
tgctggagtg tccgtcctgg gagtggctgc tgacccagtc caggagaccc atgcctgcca 120
tggtccctgg gtggaaccat ggcaacatca cccgctccaa ggcagaggag ctactttcca 180
gagccggcaa ggacgggagc ttccttgtgc gtgccagcga gtccatcccc cgggcctacg 240
cactctgcgt gctgtgagta cccgtctcct cccaactgtc agatccaggg accactgagg 300
tgtggatcca aaggggaac ccctgtaatg ggagtttgag ttaggtttat gtcataggat 360
ggtgggacgt gactggcact tcgttgccct gtggggaggg gagaaggggg ggcagcatct 420
gaggcccact ttggaccttg gcgttcgagt tcagggagcc tgtgtcatga caggcttgtg 480
tggtgctagg gctctctgag tgactctggg cctcccctat actgcagccc tccatgacct 540
gtggtgccag gggtctccgt gagttcctgt ggcaggagca gggacagagt caggaagaaa 600
ctcaggcctc tctggtggag ggtgtattgg aatgcatttt ggtcagctca agcgtcagtc 660
agggactcag tgaagggcaa ccttggcaaa agggtcccct cctccaccct gctagtatgt 720
ggtcttagag ctaattctat ttggggagct gagtccgggg tgcatttaat caggatagga 780
ttcctcgtgt acacatttta ctctaggctg caaggacaca ggaagccaca gagctgctct 840
ctgagtaggt ctctgtccct ccttgcactc agctatgtcc ctccctatcc aggtctctgc 900
ccccgttggt acccccccac agcccagtgt gaagatgttg ctgaatatgc tggttatcct 960
aacaacagag gaaacagcca cttgctgaag gttccttttt aacactccgt ccgtctagcg 1020
tcttctgagg aaggccgcca cccctatagt cctgtggtca gaccctgtcc aggcttcagg 1080
ctggagcagg gcaggaacac tgtcaggaag ggtgtgccta cttgaacagc acaggtacca 1140
tctgatagat tgtccctgga ctgagagaaa gatctttcag agcagctagc tgccccccc 1200
ccaatcttca tgcaggaggg aagtgggtga ctgaccacag actgttctga gctctgactc 1260
atgttgggac ctgctgtgct aggcatttgt tggctagttt catcttcaag agagccagga 1320
gtgtgggttc tatgaacacc tagacctgcg attaaggacc atgagggctg gacaggttac 1380
aaacatcgag gataaggtcg gtcatttgct cccagagacc actagcttgt gctgcctacc 1440
tcctaccacc atcaggctag gcatgccagt ggacttgaat ggaagtaaga ggagctagag 1500
tgttcacaga gttgggggca ggggcactag aagcccgtac aggctcacgg ctggctgagc 1560
atcttgtgct ggtggagtta gcagccagct tcctgcacac ccaccagcat atttcaggag 1620
aagcactgca ggtggcctgg acctccccaa agcactgtct cttggggaat ctaccagtga 1680
gaggccctga aaggaaagag ggcaggaagg tacttttcag tgtgtcacaa gctcagcctg 1740

```
actctactaa cccagttata ttttttttctg tcccagtggg atttgggcca agggcaattc    1800
tgttgggagt tggctatggg ctgagagact gttgtatgca ttggttacaa atgcacacca    1860
cggtatgggt tctcttgtgc taaactggca gcatctggaa ggctggagtc agagcagaag    1920
gagcctgagc caagaccaag ggatcttgga ggaccttcgg gcaacacaga ccttgccttt    1980
tttcttcatc tgactccccc gcctcagctg tcttaagtcc aagcaaagca aagatgacac    2040
ggagattttc aaactaaagg aatgactacc acaacctcag tgttctataa tgacccccc    2100
cacacacaca cacaccctaa aaatgtagga aaggcaggac agtggtggtg cacaccttta    2160
atcccagcac tcgggaggca gaggcaggca gatttctgag ttcgaggcca gcctgatcta    2220
caaagtgagc tccaggacag ccagggctac acagagaaac cctgtctcga aaaaaaacaa    2280
aaaaacaaaa aatgtaggac aggcatactt tttaatttga aaaattgtta gaagcctgcc    2340
ttttcatgca aagagactta acttcctgaa aaaaacaaac tttagatcct tactttctcc    2400
tgttctctgt ggatgatgga acctcccccа cctcatcgct gccccctcgc catctgccct    2460
ggccagagtc caggctcctg cccaagaaga taagtcagca gcttgtagga cagcaacaag    2520
gtcgaggtca gagatggggc tgtgaaggag agatggggca tggggtacat gtgggataca    2580
gggcagctga gttctctttg gtttcaggag tgatagattt ggcacgtgtg gtgtcttgct    2640
ggacatcagc cagtctgtgt gggctctggg gcaggcagct gtgggtcagg tgccgctggg    2700
actccatcca tgttccttct tgtgaccgaa gggacaccaa caggggccca gtgcatgtgt    2760
tgggtttgtt tggccctctg ggaagaatgc agcattgtga ggagaacctt cctgctctga    2820
gatttgacta gacatgacat gggagaggga ggtaaatgct taaagacaag gttgcaattc    2880
agttccaccc acgtgacacg agtgcacatt caccccacac attgaccttt gtttccttca    2940
gagagatcaa tcctgtcact tatcaccaag caagaaggct tttctttctt ctcagtcatc    3000
attcagagag ccttggttta ggggagtctg gagttacact ggggccccgg agcactggcc    3060
tggggagacc ttgctagtat gactggagtt ctgttctacc ttccttgaaa gggaatgtgt    3120
gccttttgag tggggcctgt atcaccttca cttgagtaga gcctgtgtca ccttcactca    3180
tttgtactca gttcctccgt gatgtcagct cccttcccag gagcctgtgc accctgttgt    3240
ggggtattag ccaggtggat ggagacctat taagagtctc atgagcaggg acagcgcagc    3300
tacaccatgt gttgcagaga agacaatgct tctgagggta gctcagcaga tgccgggttt    3360
gtgggtcatc ccagtagctt cttattgctg aagctacata gcaagaattt gaatgatgac    3420
ccagcacttg gaaacaactt gctctttcta aaagagatga caaggccaga ttcaacttgg    3480
tcaagatgac tgttgtctat gtgaatggca tttcccctaa actactctgg agtgcttcct    3540
cccttgcagg gagaatatgg ttgcctctgg tccagccccg atgaaggatt ccctaagcaa    3600
gtggtcttcc atagtgcacc caggcctggt ggtggggtaa gctctgtccc agggatagaa    3660
tgccaatagc cttggtagct ctggcagtgc agaaagaaag gagaaaaggc atgggacatt    3720
tacaaccaaa actgccctca gagaaggcat tctagtcttt tgaaagaaac ggtgtgacca    3780
gacactgggt gtgataagcc tgccagggga gataaaaaca ggccctgctt ttaggactac    3840
tggagagcag ggtgaagatc acacacttat actgtctcac acttgttctt tggtaggaag    3900
agaactgcag agaaggagtt ggtgagggtg aaaatgccta gcagagggag tcagggcatg    3960
aggtcatctc ccctccccat cctccattga aaatgtctat aaggtttcca cagcatgata    4020
atggcttttg tgcaaacaca gtgtggcact gttttccata tctggcatca aaggcaaatg    4080
agggaaatac ctgcataggc agaacccaga gctgaaggcc tatgggctcc tgagaaacat    4140
gagaaaaggc ctttgtttga gaaggatgct cttgaagact cattgtgtcc aagaccagga    4200
ggaagggctg aacccaggag ggcctattta aggcctattg tatcatttat gtaagtggca    4260
gagtgactat gtttctgccc atacccagta ccctggagct gtcttcccag gtacagggta    4320
ggcactggct taggcgcata ggttaaattc acctacaacg caaggccatt agcacttctt    4380
aacgatccca tctctctgcc tgccacagag atggaacaag aactgctatc gttacccaat    4440
tcatgccagg ccacgtagtc ccaaggagca agactctgga gctgcctgga atctcttgtg    4500
tcatgtcagc aagacaccta gaaccccagc tccaaagaga ggctggccag accagttcac    4560
tggctttaca gtgcctcagc tgaggttaag gtacccactg ttaagtcacc catccacatt    4620
ctagtttgtg gttcaatggt ccttagtaat ggtcacagag ttacgcaact gacaccacca    4680
tctaatttca gagagttctc attattccca gaagaactgc ccccccccc cacacacaca    4740
catgtgagca gctgcttgct tttcttctga cctctggaaa gagccagtct actttctgcc    4800
tttaaggatt tgcctgactc ttgcattttg tgtatgtgaa attatatggc gtgcagcttt    4860
tgacatctgg atctgtccac ttagcacaat gccctatgct cattgaattg tagcaggcat    4920
ccaatggttt gataacccac tgtgtggaca caccacattc cccggttgag tggcgttttg    4980
cactgttctt actttttgac tgttttgaac aacagttgct gtgaacattc atcaaccagg    5040
ctctgtgtgg atgtctgttt gcaggagtct tggggagtga gaggcagggc tgggtcatgt    5100
ggtgattttt atgtttagcg ttttgaggaa gtactgaact cttgtgctca gctagagctc    5160
```

taaccagaca ttgtgagggt ccagtttctc ttctagttac tctctctctc tctctctctc 5220
tctctctctc tctctctctc tctgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg 5280
tgtgtgtgtg catgtgtaca tttgcacact gtgtgtgtgt cagccaaaag acaacttcag 5340
gagttagttc gctccttcta ccatgtgggt cccagggact gaactcagat agtcaggctt 5400
aatgacaagt acccttaccc tctgagctac ctcactggcc ctacctttt catttcttta 5460
tttaagacag aggctctcaa tgccctggaa cttacgtagg ttaggtgggc tgatgttaga 5520
agcaccaagc ccagttagtg gtggtcggtt gcttgttaat gtagtctggg catcaaactc 5580
aggtgcttat gctttcaagg gaagcacttt atcagctctt acctccccag cccctcactt 5640
gtttgtgtgt ttgtgagcta tggtctcaca tagcctaagc cagcctcaaa ctccctattg 5700
tagtcaaaat tggctttgaa ctcttgatca tcctgcctcc actttctaaa tgtattcacc 5760
acaatatctg gctgtctttt tattctaccc agcttagggg ctgcaaagca gcacagcatc 5820
taactgtggt tttagtttgt atccctaatg ttaataatgc taatgggtca tttgtatgtc 5880
ttacttggag aactgtctgt atagtctttg cacattgaaa tagctatatc atttcaatga 5940
taaaatagga agacaggagt aatgtggata ttccatagcc taacttgaaa cctctaggtg 6000
ctttgcatat gtaaaaatag agctggccat tagttttag agctgaaagc aaccagtgat 6060
atctcagaag cagaagaccc atctgtggag aactttccat gacaggagag gagagctgtg 6120
acagtgtcac ttccgggact tcctggaggc ctctgggaga cacagagctg tcatgtgggg 6180
cctccacaga ggaagtgctc aaagtgactg aggcaggaag aggacttcaa tgacgcaaga 6240
gtgtctggtg gctgctgagc cacaggggtt gacagcctgg aagcctgggg agggaggagc 6300
ctgggtacag acagcaagaa atgcttagag gactgggtat gaagatgaag tctaggagag 6360
aggctggccc ccaccacctt ctgacctgag ctccaactta gtaaagagat gccgaaggga 6420
tgaagtcctc tgatcgctaa aatgctagct gttctatggg aggaaacacc atgttggtgg 6480
gcacttgccc tttgagagag agggtgacgg aggtggtgag ttcaatctcc acagcccaca 6540
tggtgcatgt ttgtaagccc agttctagga aggtggagac cggagggtcc ctagggttca 6600
tgggccagcc agactagctg aattagtgag actcagacca tgtctcaaaa caaacaaaca 6660
ccaaaacaac aacaacaaac cctagagggt atctgaagga ggacatcatg attgtcctct 6720
gatctctact cacacattca tgtatgcaca catgtgccct tacacacata catgggggca 6780
tactgttaac atgtgccagc acccatcaat ttgggtttac tttgcaattg aagtaacact 6840
gtgaagggct ggccaaagga acctctgaaa aaagagagtg cccacgtggc cctgcagctg 6900
gaataggcag tgtagaggtg gacagacctg gtatgaaaca gcaaagtctt ctttcaggtt 6960
aacagtagta ttggctgtgc ggtggtacac acaggtgact gcacactccg aaagctgagc 7020
aggaagattg ctgaaagttt cagaccagct tcccaggcta catagtgagg ccctatctca 7080
acttaaccct cacagtgagt taagcataaa gtaaaaagtg taaaataaac aaaaattgtc 7140
acccaatcac taagagacta gtgcaagcca tgtgaatttg ttgatgtatc ttctcatacc 7200
catttctatc acataatata catacacaca catacacaca catacacact cacatatatg 7260
tacacaccac ctttcatttg tggtctgggg tttccaacaa catcaatttt tcctgcctag 7320
ttagaatttc ttccttaaaa tcactcttgc tggaagccta ggcttctctt ccagtatgtt 7380
tatgctttat ggaatgattc tttgattgac atttacatta ccagcaatct tgctgtaagt 7440
gtggcagcag caaaaaaggt cttccagcag aattgctaca tatgttactg gttacaaatc 7500
cccactgcag ttagagcacc agcaggaatc gtcttcaaag actggacaca aaatggcaac 7560
ttgctttcca gaaaaattga cttaatgtat gcttctggca agcagcatga gaaaatgccc 7620
acttccttcc aatatcctat tctacggaat gttacttta agaaaattgg agtcataaca 7680
tgaaaggtca tctgtgatca ttttgttttg cttttgattg ctgatcctgt ttgagaggtt 7740
tgctggccat tttcttttgt ggatggtcct ctattctctt agcttctttt cttcaattaa 7800
gactgtgtgt agcttgggaa gaagggttgt gataagtccc aggctaccac ggactgtata 7860
atgaggtctt atcttaatag agggtggagt ggagctggag atatggccca gtggttaagt 7920
gtttgcttag tgtgtgtgag gtcctaggtt caatccccag taccaaggaa tgctgcaagc 7980
tttgccatct gcctatgtga gttctgtcat gtttgtgaca aatatatctt cacgatggtc 8040
tgcctttaag gtttagattc tctgacatgg gagggtcact gttttctact gggacttctt 8100
cctgggattt ctaacacaga gactccttct gcacacagaa atccactcag aaaggttaag 8160
actgggtctt tgtagctcat ccacttgggt taagggagga acctccagca gtttatcgaa 8220
ttacactggg ccccagcctt gtcatctgaa aaccaagtat aagaggatct atgtggtaga 8280
attgttgcat gtgttgatgt gtgtatctgt atcatgtatc atgatactaa caaatgtagc 8340
caccgtggct attattttta tatgttttat tgttcactca atttagcttc aatagctaaa 8400
tcttcttgga cttaatagat gacaagtgat gaagatttat gtgtattttt ctcaagtaac 8460
catccttccc caacctttta ttaaatgact tatgcttttt ccctccatga ttcctttgtc 8520
taccatatgc taagctcttt ctgggttgtt tatctgtcca agaataccca actacaccac 8580

```
tacctttaca aagacataag cataattttt caaaatgttt tggggttatt atctcttata    8640
ttcaagtgtt ttcccaggta aaccagatag taatattttt caacttggac aaaacttact    8700
gttatgactt tcactgggat ggcaggaaat cctaatgtaa gctgtgagct tgacagtctt    8760
attgtgctcc ctctctgatt gtgtaatata tttcacacag accctgagtg tttcttgtct    8820
gggtctctgc tgatgacttg tatgtttttg ttattgttgt gaatggcaca tttctcccgc    8880
agttactggt tcatgataag cggttgcctg catagaattt tatgctttca gctttcattt    8940
tgtcccactc tttggccact cctcattcat ctttcagctg cttctcctgt cttttctttt    9000
taaagagatc ttgttatctg tggtgtgtgt gataccatcc atgagcacag atgtgtgcgc    9060
acacatgttt gtgtttgcct gcacatgggc caggaggtga gatgaggaca cctgaaatca    9120
tgatccatca ctctctgcct tgttcacttg agacaatgct tctctctaag catggggctt    9180
gctgattttg gctatgctga ctgaccaggt ccagcaagcc tcctgtctct gctacccacc    9240
caccccacct cagcactgga tttatagatg tgtgtgacca cagctcagtt ttcttaaatt    9300
ttcagacagg acctcactat gtatccttgg ctggtttaga aatcacaaag atgggcctac    9360
ttcccaagat ccagtattca aggcacacac caccacacag actctctgtc ctgctttttc    9420
acgtgggctc cggggatcta aactcaggtc tttatgctca cacagtgagc actcttactc    9480
actgaacgtt ctacccagcc caacttttct gaataggcct tcatatcatc tgcaaaccag    9540
gatcatagtg tttgcttcct aatgcttaca accactttct caaatgccat agcctattat    9600
attggccagt atttctaaag tggtagagtg gcaactgttg gaaaaaggaa acgtctgcct    9660
taaggagagc ttcacatacc ctgccaccgg gttgcctttt ggtcagggcg ttttattaca    9720
agaatggaat tgggacatga gggttctggg aaggcagata cctgtcaggg cctccccatg    9780
ctgttttgaa acaatgggtt tatttaagat agagatagct cctttatatc ctttgttggg    9840
ttggggccct cttttgagtg gtgataaggt ctggtgattc tttttctttt tctggcaact    9900
ggggactgag ggaaacagct tcttttagaa tcaaaagcca ttttgcccaa catatcttat    9960
gaaagagaga cagacaggca gacaggcaga aagaaagaag aaaaagaaaa aggaaaacac   10020
caagatgaat ggggatgaaa ggacccacca aaccaacaag acacatgtag tattgggcca   10080
gtggtaaact cagacagaca gacagacaga gggctccctt gtgagggcct aactcccgtg   10140
tctttgctta aagcattagg tcaagagtga ccgcttgggt ccctcctgct gaggagggtt   10200
ctctccctgt ctctccccca cctctcacct gttggccacc tcaaccctct tcctgttttc   10260
tctcttttgg aggggcccaa ggaagcaggg gatagcctct gttaatatat caagtctcag   10320
ccctgtgact ccagtgttct ctaggttgga gtagcaagaa taaacactct tgcttgctgg   10380
ggcttaggta cattttctct ggttctgtgc tagccaatat gccttgcctc cctctttatt   10440
ggtgcatagg gacagggtca ctatgtaacc caggttggct ttgaactctc tatactcctg   10500
cttcaacctc ccaaatgcta gggttgtagg tgcacaggac catacccact tccatatggc   10560
tttctgaatg caatctttcc attttgactg ttttctaccc acctggccca gaagtgtaca   10620
gcatttatcc aagaaaagag aaagcacact gcaaagaatg ttttctagaa attagaagtg   10680
acacataact agagttctgg tgggatttac aagagcagca tgtggagcga gacacagcct   10740
ggactaaagt ttggactaat gttttcataa gggtgatggt tatgatgctg tcgaatattg   10800
acagtgctca ggcagggtca tctgagatac ttctgttggg ttcagttctg attttggcct   10860
taatcgtgtg tggttgtttc tagaatatct tttaaaaggc caggcaagcc gggcgtggtg   10920
gtgcacacct ttaatcccag catttgggag gcagaggcag gtgaatttct gagttcaagg   10980
ccagcctggt ctacagagta agttccagga cagccagggc tacacagaga aaccctgtct   11040
caaaaaacaa aaaaataaaa ataaaaataa aaggccaggt gggctttttt tcctacactt   11100
ggttaaattt tccttttttgg gtcagccata tctttatctc tctgtccttt ctttctgtct   11160
gggagcccat gtgtttcacc ggctccttag ggctattcca taagtagctg ccctcggcta   11220
aggaaagcca ttctcacgct ccctcctttg cttaggagga agggtcagag ctcctgagag   11280
ataaggccag gtcagatcag ggcagagatg gctctacaat tcccctgcct ggctctgtac   11340
gggttgagcc taggagccat aggggcgtat aagtgggcgt ggtgaggatt ctctgtgttc   11400
cagtgatatc tctgagaggc gagatgctat cgggccagat ggtggtgata tcgacctggc   11460
ctgtgaaatg ggtggggttg gggtggaaca ttaagttgtg actagagcca gaagagcctc   11520
tcaagagaag ctttcccatg gggaatgaag caacgaagtc atgaactctc gaaaaagcat   11580
tgccccacat ttactttaag tatatatgaa ttcaaaacag taaaagatca taaagctggc   11640
tacatagaaa aatataaatg tatacagcaa cggaagggaa ccaacaaaca ttgttgaaat   11700
gaatatcaaa atccacagta gcatagatca gagtcttgtc tgataataag cccaggatga   11760
tgtttaaaaa aaaaacaaaa aacaaaaagc aaagagatgg aaagggattc aaagaaaaaa   11820
aaaacagaaa cagccaagaa caaggacaaa ccaaaccttt ctcgtgggga gaggcaaatt   11880
aaggcaagct gccatagttc tccccctctc agatctgaaa agagatctga taagccagtc   11940
agcaggctgt ggggaagtag acactctcat atctgttagg aggaaggaaa aaaaacccga   12000
```

cagactcact ggagcgcagc tgtcagtctt tatcaaaatg taaaatgcca gcaagacatg 12060
atggctcata cctgacatcc tagcacttgg gagactgagg caggaggatt atcttgtgtc 12120
agaagccgac ctaggctaca taaaagtgtg cgtgtccacg caagcatgtt tgtaaagtgc 12180
ctatgccttt ttgtctggct gactctgtct agttcattgt ttatccacat aactagcaac 12240
ccctgacacg tgtggtgtcc agtgtgagcc aggcgctgtt ccaaaggctt tcgatacatg 12300
gatgctttta gtcttcccag caacgaatag atatggcatt gctgtcatct ccattttaca 12360
atgaagaaac aaaaacaaag acttcagcca tgtatgcatg ggttaaccat gtatgacaca 12420
cggatgtagc catgtataca cacacagatg ttgccatgta gccacatata catggatgtt 12480
tgttcaccac aacattctta cacaggcaaa acattcaaaa taactaccac caccaggctc 12540
tggtcaaggg catgacaata catgtaatgt ggtagaatgc tgagaacatg agcaaaccat 12600
ctgtcatgtg tgtcttggtt acttccctat tgctatgaag agacaccatg accaaggcaa 12660
tgtataaaag gatatcttgg aattcatgat catccttagg cttagaatcc atgattatca 12720
gggcagggaa cacggaagca ggcaggtagc tgagaactta catcctgatc cttttacaac 12780
ctcaatgtct acccccagtc acacacctcc ccaacaaggc caacctccta atctttccca 12840
gttttaccaa ctgggaacca agtgtttgaa cataggagcc tatgagaaac attcgcattc 12900
aaaccaccac aatgtggaag actttgatga tattttatga agaaagcaaa ctttagaata 12960
acataaagag taagatttta tatatttaaa aaatttaaag atatatacac acacatatat 13020
acatatatac acatacactc acaatctttt atgcatgtgt tttaaaatat tttataaata 13080
aggataagtc taagaaagtc atggatgata ggagatataa tttttttcat tctatatctt 13140
ctgtgtgtgg tgccttttt caatatataa tctttttttg tttgttttgt gagacagtct 13200
cacatagccc aggctggcct ccaaccctct atgtaactga ggctaacttt gaacttttga 13260
ttctcttgcc tctacctcct gagtactggg attattggtg tatgtcgcca ttcctggctt 13320
atgatgggaa ctgagcttag ggcttcgtga atgttaggca aacactacca accaagcaaa 13380
tgcacccagc cctcgatata aaacttaaac tgagccaggc gtggcgcacg cctttaatcc 13440
cagcactcag gaggcagagg caggtggatt tctgagttcg aggccagcct ggtctacaaa 13500
gtgagttcca ggacagccag ggttatagag aaaccctgtc tcaaaaaaaa caaaaaacaa 13560
aacaaaacaa aacaaaactt aaactgagtt taatagtaac tgtgaaattg taagttaata 13620
gtaaatgtga aattgtagtg ttgtgtggca tgtggaaggg attagtgaag tgacttttta 13680
tgaggttgat ttttatatca tgggtgtatt tcctgcatat atctatgtgc actctgtgca 13740
tgcagtgtct gtggagacca gaagaggatg ctggattcct tggaactgga attaaagttg 13800
ggtgcaggcc tccatgtggg tgctaggaac ggaacccagg tcctcctaga agatcatcaa 13860
gtgctcttaa ctactgacat ttctctccag cctccctaac tctcatttct agattagaag 13920
cattgttgat agagttttac agcccccact gcctttgggc aataggagta cctagctttc 13980
actgttcatg agaaccccaa atagaggaag acgggtaaga ggccattcta gcaggattaa 14040
ggaagacctg gtgatggtga tgcttcccaa aactgctcac agctcagctg tggggccact 14100
gaactcctga ggccaccaac agacccccat ggccaccatg ctaccagata ccagatggct 14160
ggtgactggt cacctacggc cccttcccag agaatcctta attttaataa gttcttcctt 14220
tcagagattc taggaaagtt aaaaaaaaaa aaaaaggtgg gggaggggggg ctgggattct 14280
taaaccaaaa atgtagaagg agctcttggc tggttggcac ttcctctata gaaacaacca 14340
gagctctcca acccagtgtc tgaaaggggt aggactcatt gcaggaagaa ggaagccatt 14400
aaaacaagga gtccaggagg agccaagtag ggggaatgag gtctcagcag aaggtgggta 14460
ctggggagcc tcaccctggc cgatgtgctc cttccgcagt tgtagctgag agctggtcat 14520
catgatacgg atcttgctga gctcctggga gttttggctc ttcacacctg tctcttgctc 14580
cattctcttt gatgatgtgg tagctgaggg gttgaggaag cccctctaga tgatggagca 14640
tttagatcag accacatagc tatctagcag tgacgtaaat ggcatatcca tggccactct 14700
tggtgtgctg aggggcagga acgctgtgct gtatgcagga ctgtgaggca tcaggttcag 14760
ctctgcctca ccatcatccc tcagaccacc gtcatccccg acatttgcag gctttgtacg 14820
cagactcccc cccaagctgt gtcagacagc cctgttctat ctggctctgt gttttcctgc 14880
atggtttgta atgccgtctt tgatttattt caggttccgg aattgtgttt acacttacag 14940
gattctgccc aatgaggacg ataaattcac tgttcaggta agtcccaaac caacttgtgc 15000
tttccctggg agaaaagaaa aaagagaaag gaaaaacagg tattctgaat agttgcaatg 15060
tgagaaccta gtagagctta gctttgtccc ccaaagctat gtgtctgggg aactgagtga 15120
ataacagtga ctcctcatag atgacagtta taagatgata gtcatgcatg ctaaggatta 15180
ttggtggcca tggctgagaa aaggacacag tggtgaagga aagccaagcc agtgcctatg 15240
ccttgattta ttcttttgtt tgacttccat ggatgagcag atgtcagtgg cccaggaagg 15300
agcagaggga ggtgatcttg caagagagat gggagattag atgagatggg gcctttaaat 15360
tcccatagct cccttctgcc caaagcattg gtcatggcat ctgaatcact gtaaaaaggc 15420

caaatataca aagccataaa taaacttgtt gattgattga ttgattgatt gattgattga 15480
tgagttatgc cacagtatgt gtgtgtagag gtcaaaggac aacttgaagg agttggttct 15540
ctctttctgc catgtgggtt ctggggattg aactcaaatc ctaaggcttg ggtggttact 15600
gcctttaccc actgaggcat ctcaccagcc acagctggat tttaaaaggg aaacaaacaa 15660
acaaacaaac aaacagacaa acaaacaaaa aatcatcagt gccaatcaaa gccggggaag 15720
accacccttg cagtgtttgt gtgacaaaca cagggacttg tctgaccaca gagtgtggca 15780
gctgagtgat gtggaagcca gagaagctag tttttgcctg acttcattgg ggaagcatgg 15840
gggcgtgcag aaaggaggtg acagccccca tgtcctcagt agaggagagg gctgtgtgtc 15900
attaagctca cagcgaccat agtaagagtc acaacagcag atgggacatc tgtgtgagca 15960
gtgcagtaag gaaagacaag catctaggga ctgggtcacc agaacaacaa ggagcaaagt 16020
cttcaggatt ttgctgaaaa gaaaggggtg ggaagggaga tgtgaggcat gttggtggct 16080
tccggaatgt gacaggaatg atgaattgtt tcgatcaaag gggtagactc agagctgcct 16140
tgttgaaagg cagtccttag gcaaaaggcc ccagaaacat ggaagcgaga gagaagttat 16200
gaacagaggc aaagcctcgg agggcagctc atgatggaat tcaaaagaaa ccacatcagt 16260
acttccaaca gtcacttcct ccagacggtg ggactcgaag ggggccagtt ccatccaggg 16320
tgtggcaaag cttccgtcta ggggaagggt ctgctcatac tgagcttcat gggagaggcc 16380
agcgtgaacc actgtgggct ggagtaacgg ctgtactgca tagcatggtg gtcaggctga 16440
gcacagttca tacttttttct cgtgtgcgca tgaacaccag gatggaaggc ggggcggatg 16500
ggaggcaggg cggatggaag gtgtggtcta ggggtcaggg agtccctttc ccagcccagt 16560
cattggtggg aagcaacaga acaaatctga gtacccgcaa gcaagcgccc ttttttaagc 16620
ttggcaggtg ttcgttgcct gaatactttc tgagatgtgg acgccaccag cttctggaga 16680
catggctcag gggaggtggc ttaggcatgc agccgtggtc tctacttttt ttttttttt 16740
tttttttct aagcctagga ctgtgtgaac tttgcctggt ggcttggtgt ccttggcaag 16800
tggcttcctg ctaggagtca actctcttaa gttaaaagtt agtcggcatc tttgatcttt 16860
ctccccttt gtggtttctt ctgttccttc aaagtgatga agactttaag tatgactgga 16920
aaagaggtga ctctgcggct tcttgttgag ggtttgtgcc cacagcttgt gaacccataa 16980
ctctgcctgt ctggagtctg tacccacgtt ctcttcactc ttgcttccgt ggcttcctgt 17040
gaccccagaa atgaccacag accttcagtg ggactaggag cttcagtgtg ctttgggagt 17100
ccctctggct gatttttttt ttttttttt tgcaaaggtt tagtgccttt ggtggcctga 17160
aggacactag agaagcccag gccttgggac cagaggctct ggggccctgg ctgggttcta 17220
cctcagtgag tcagctcttc tctgtgggtt gctcctcccc tgctgacagc actgctcttc 17280
tcagatccca gagggaacct gattggctca gccaccaaca gctctctctg cctgaaccac 17340
tcaccagcta ctgagctctt tgctgcttgg ctgttaggaa gcaggtacct gcccaccact 17400
gacagatgag gacagatcat agcctcctgt ggctacctca gccagggcag tgagcaggca 17460
gtttttatag aagagatatg gccttgctgg acacaattgg tttcagtctt gcacttgacc 17520
tatgaccttc ttggtcatct ataccagcca agggacagga agccatgggc ctaaatggtt 17580
tgaaagtccc cacctataag gatgtcttgt cctataaaaa acaaacaagc aaacaaaaca 17640
acaacaacaa aatccccaca catttacaag tgctggagat gtagataagc tcaattggta 17700
gagagcctat ctagcatgtg cagacatggt tcagttccct agtactgtgt aacacctggc 17760
atagtagcac atccctgtaa tcttagcact tggggataga ggtatagagt taagaggatg 17820
gaaagtcttc aggtacatag tgagttagag atcggcacac atgcatgtgc acacgctcat 17880
gcacacacac atgcatgcac atgtacagaa cacacaccag acatgctaaa acaaacaata 17940
taaggaacaa atcaaagctc acagccagtg aaggcaaggc agaaacattc ttcaggaagc 18000
gttggtattt ggccactgct tattttaact ccagttcctc ttttcatgtt gaagcaaaac 18060
ctgcattctt tcctgagctt agccacagtt tctgtataac tcagtcctta tacctaaacc 18120
tgcttagtgt cagcagcagc aggagcaagg ctgagactgc acagcgacac accaagggcc 18180
ttcatggtcc cagctcttct tatgggaggt ccttgttgaa acattcattc caccttcaat 18240
ctttcccaag gacaatctgt atctcgtggt ccagaaggct cactgatgga gggaagcacg 18300
acccccacca ctcagggact cagctctgac ttcattactt atagacctac aaaatcagag 18360
gtgggaaggg tggcagcagg attctcttgt ttaacgcggt agacactagc gttcacaacc 18420
ctgcttccaa gctaagtctt ctttctcagg gctcagaaca tgtttccagg cataaatggg 18480
tactgaaaca cttcggcagt atgggaccca agaccatttc acggagggga ggcagggtag 18540
tggtctctct gagatgggcc cttccttgcc tatttggatg gtatttgatc atgaatagca 18600
cttggctgtg tgtcctgtgc attcaggatt ttgcagccac atggctttgt tcagggactg 18660
tgaagagacc aaacaggaag atagaatgtt acatctcccc tgctgtggtc acaattctgt 18720
tcacgaatga cagggatggt ctccatggag gccattgttc tctttactgc ccttgttctt 18780
agtaaggaga catccatgtc cagttcttgg taagaaggga gacaggggac tggtgagatg 18840

gctcaacggt taagagtgcc gactgctctt ccaaaggccc tgagttcaaa tcccagcaac 18900
cacatggtgg ctcacaacca tctgtaacga aatctgatgc cctcttctgg agtgtctgaa 18960
gacagctaca gtgcacttac atataataaa taaataaata tttaaaaaaa aaaaaaaaaa 19020
aagaagggag acagccacat gttcttcagt ctgcaaggtg aaaaaagaat gggaagaaaa 19080
tgggagggaa agagatgctg agaacagtat aaaggggggg gactgaaaga cataaaagaa 19140
aaggaataat gagatgaggg aaagataatg aaggaaggaa cagtcaagaa agaggaaggc 19200
tgtggacgtg atgggtacgg tttctacaaa ggaaaagagg tgggtgcatg cagggcgaga 19260
agctgcaaag gtgatggagg atgaagatgg aagtgtgctg tgggctgtgg aaggtgtctg 19320
tggaggatgt ctatccctgg ctctgctccg ggcactgaag caaggatcca cagacactgt 19380
gagacaggtt taacgcagat tctcagaccc agagagtgtt ctgatttaaa gtccatgcgg 19440
acatcctctg ggtactgtgt gaggccagag gcaggcaggc aggatgctta cttagcccat 19500
cgggccccgg gttcccatgc ctagaggcgg tgttcgtgtg aatcctgggt ccttcctgaa 19560
tctacaaagg aagtgactct cccctcaggg aagtgacatc catcttgggc cgctgacttc 19620
agagtgagct gttcacccag ttgttgcctt ttggagtagc tttgctccct gggaagccag 19680
ggagggtctg aggcaggaca agaagtctta acagacaact gtgttgtgtc tgaggaggct 19740
gaagagctct gagcagcctt tagacatgat acctctagcc aggtccacca ctgaagcgaa 19800
taggggcctc tgggcttctc ttgctgcctg acacttccta tgaagaagaa aaaagcccag 19860
ctcatgattc tgtcgcataa atatctctgt gaacttcctg tctctgctga gccggcagtg 19920
ttggggataa ggactgtgaa aactggggta ctgtgggttt tcatgcacgt tcagaagaag 19980
agaggctttg agatggtctt tcccacagac gccaaggtta ctcttccaga cagaaggaaa 20040
ccccactagc cagagcctgt cgtgatctgt gtagatcaga cgcgtgactc aagttcctgg 20100
tccctggtga tgtctgtcac attgctctgg gcagttcggc atctggagaa ggaacaaaac 20160
cacgttttca gagctggccg gtcacaagcg tcggggaggg cgaacggttg agccgttgtt 20220
atgaagttca ggaagcgggt ggcagctgtt attgttctaa gtggtcacca agtgaaataa 20280
agattccgtc tgagcctctg ggcaaaggat gtggctcaaa cacacatttc ctgaggttgt 20340
acaaggaaga atgcagaaca ggctcagaag ctagcatggc caatgtgaag acagggagac 20400
ccctttgtgt cctggcaacc atattctggg gccctgtgtt ttaactttgc tttattttat 20460
tttattttat ctgctgtaga tataggtcta ttttctaagc ctctgactat ctcaggacat 20520
cttgcagctg ggttcatatt taaaagggaa attcttgggg ccaagtttcc cttcagcatc 20580
ccccagaggt gtcgttccat agcctttgaa aatgtagtat gaataagtcc caagcaccaa 20640
cccaaggcca gctcgtgcca gttgtagtag agagacaggt attacctttt gctcgctgca 20700
tcctttcctt ctgtggatcc tgctgagatt gatcaatttg caacacaggt ctttcaagac 20760
tcaaaggttt ttcttgtttg actttttttt ttttttaag ccagtttatc tgttttattt 20820
ttctcagatt ctactgagaa agcccagtcc tgccctcatt tgccagcctt ctggaacact 20880
cttactccca gaaacaaaca agcaaacaaa caagcacccc tcacctgccc cttctcctgt 20940
gcgcctggag ggagagctaa ttgattctca ggtgtctggg ggtaggggtg tttacgctct 21000
gccatctctg cctctaaatg gcttatcctt ttctccccat ctctgtgcac ccaagccgtc 21060
ttgttgctaa ttctccatat cccagaatgt tatggtctgc tttgtttgtg ttttcttctc 21120
ctgcatagca cagagctgac actaactgtt gataggctaa gcccagggaa cagtactcaa 21180
ttgcagaggg gaaggactcc gtaggggcgt cttatcaaaa agtctttctt agaaacagac 21240
acaagacaaa atggctccct ccctcccaac tagaatatga tcattggaat tgctgctgct 21300
gtttcacaac cgtgagaaaa ggtgactgga gaatgggacc atgcagacag cagagcagaa 21360
agataaaatg aagccaggtc ttgaatggtg ctattggatt agctgcatcg aattaacaag 21420
tctaggagct ggcctgccca tgggcttcct gttagtgaga tgactctttt gtgttcaagt 21480
acatatagga ttctgaaaca cgtagctgaa agtggtccaa catacaaggt ttggtctatg 21540
ctaggggggа atctcaaaaa aggaaaaccg tacacttcag attgagcacc gagagtcact 21600
gcacaggtcc acaggctgtg gtctgcaacc tcgatcacac ctctctcctt agtcctggcc 21660
agtgcatgct gccgggattg gttctgtacc atccagacat ccctaacgat ttgtctctgg 21720
actcaatatc ctattgctct cactcaaatc agaaaaagta caaaaatcaa tcagcatgct 21780
agcgacctct gtgccctgga aagatgtctc catttctgtt tctccatggg atgatttagt 21840
ttctcatttc agctttgccc tttgagagca ggtgtcattc agagttgtat tgagcagatg 21900
agagtctttc ctcggaggaa ataccatttc ccccagtgga accttgctgg atcactcata 21960
acatagccga gggcgtgggt aaggtggctc ccaaggaaag tgctggccac acaaacatga 22020
ggatcagagt ctagaaaaaa gaagggctgc aatcccagtg cgggtggggg gggggggggc 22080
agaaatagtt ggatcccggg agctccctgg ccggctaggc tagcccaatt ggtaagctct 22140
tggctcagtg agaccctgtc tcaaaaagta agacggaagg gctggagaga ttggtgatcc 22200
agcagctgag agcactggct gctcttcctg atgaccagaa cccacgtggt ggctcacagc 22260

catctgtatt tctaagttca aaggctccag taccctcttc ttgttgcctt gctagggagc 22320
aagctcacaa agcacataga cacatgcagg caaaacgccc atccgcactt aaaaattcaa 22380
ataaaagaaa atgtaagtgg acgataactg aagtcaccct gcatttgcct ctggcctcta 22440
catacgcatg cacacatgtg tatgcgcaca cacacacaca cacacacaca cacacacaca 22500
tgcacacaca agaacagact cctcaccctc cacctgccag gtgattattc aattttatgc 22560
catggtcctt atttatttt tgaggttctg aaacccttct cagcccacag tgagaatcca 22620
gtagctatca atggattcgg tctgcttgca tcttgagaag acagaagagc tgcgggtccc 22680
actggacagc aggatttgct ctgccctgct gcagttgtct ccacagtgct gttactggaa 22740
tccatgcctt ctggcacacg cttttctccc ggtggtcaag ttcttccgtg ccctgagaag 22800
acttctctcc ccagtgactc actcttgctg gaactacctc ctgcatttca gtactgagat 22860
tccaactatg ccaccaagct aacctcagtg cctagtcact ctctactgac cctgttacca 22920
gcagctcttt catctttcct tcctgtgggt tctctctcct gggtcctcac gacctcatgg 22980
ctcaaggcta tggggtcctg tccaggtggt tactcccgaa gccatgtttt cagagaaggc 23040
tcttagccag gtcttcggat cctagcaccc aggatgctgt gtgcgttctc ttaaagatgc 23100
catctgtgct tctgaacctg tctcctcctc aggagtgctg acgcactgtg gcatggcttt 23160
cttacagagg tttaaggact gaaagaggtg gccaaataag cacagagtat agcataccac 23220
tgtaaccagt gcatgatggg gcaaataaac cagaagtcca gaggctggtt agctgccacg 23280
ccaacacagg gtgatccagg caaggctggg gacaccagag gcagaaaaga gcgagacagt 23340
ggtgagactt aagaaccacc aaagtctctg gagtttggag gaaggcagct ggaatgccta 23400
aggtggaact tgaaccaaga ttgagtatcc agggggtcct gactggcttc caatcttggc 23460
cacgaacaaa gtgggagacc ctagattctg tatctaggaa aaattgctgg cacctgctca 23520
ctggggcgga gtccttctac tttggaagaa ccaggaaggc agaggcaggg gatagggtca 23580
tggctggtac ctggaccact ggccatgtct aagagagacc tgaggacata ggggtcagag 23640
agggtcccat gacttcacct agggtcttgt tgccattgag tctgtttagt gagcctccat 23700
ggaggcaggg acttgggata cactatttca tttaatcctg ggaagaggaa tggtactcta 23760
atggaaacaa gaccagtgag atggctctcc agacccacct gtggctgacc ccgccacccc 23820
caatgtcgct ttcagagcct cttttccagg cctccccatg ggaaaccatg tttgcccaac 23880
acccctctcc actgtaccct gttctgtttt cctaacatcc cttagcacta tctacaacta 23940
gttgccagta gcaacataac ccagtagtat acagtcatac acattgatta tcttgtattt 24000
ttggatgcca gaaatccaaa aatggtcttc actggctgtg ctaaaatcca gatatgagca 24060
gagctctgca ccatttctga agacttcaga aggaaatctt cctctgtcct tcaagccagc 24120
aatggtggtc aatatgtctc acagtgccaa cagaatggcc ctgtgattat cctgattaat 24180
tggtatcctt ttgcatctca gagcagaaca tgagcaagcc tagctcactc atgctggatg 24240
acagcttatt cacaggttct gagatgggga catgaacatc tatagaaccg gatggggaaa 24300
tcattattct gcctcccaca gcagctatgt tcatgaccta tccattcagc tatcctacct 24360
atctccatcc tgagaccaga agatggatcc aagagaactg agaccttgcc tgtcttgtac 24420
tgaagtgcca gaaagccacc tacactcatt aggtactcaa taagtagtga gcaaatgcag 24480
gctctgcatt aggtactcag tgcagataat gtaaatacta gatatatcca ggcacccatt 24540
aggtactcaa taatacttga tacatccaga taccagttag gttctcaata aatactggat 24600
acatgaaaag gaaggagtat tttttgtgtt tccaaatagc cctgtccatg ctatttgcca 24660
aggtgccttt aggacctctc tgcacaggcg gatttctgag ttcgaggcca gcctggtcta 24720
caaagtgagt tccaggacag ccaagactat acagagaaac cctgtctcga aaaaccaaaa 24780
aaaaaaaaaa aaaaaaaaaa aaaataggac ctctctgcag cagccagtaa gactcgaatg 24840
gcctgggatg atgccagggt cgtggagcat ctggagcaag gctaatcaga cagttgccct 24900
gggctgcccc ttgaggcagc tcctctctgg gctggctccc agggccaagc acatccactt 24960
gtctgcctgc tgtcactcct gtctcccaga gtaacttgag ttttacggaa agtaaagaga 25020
gaagctagaa gggagggcca ctcggggatg taattaagga taaaaatatg ccatgtcctg 25080
ggctctggca acatggaaat ctagattgct catgttccca ggactgtgtt gtctgtatcc 25140
aagtggtccc aaatgctgtg tctaaattct tagcagggac catttaaaag ttatgtgtgt 25200
gtctgcctcc ttccttttct ttttcttggg atgccaagaa ttcaacccaa ggccttacat 25260
accctaagtg agtagtactc taccactgac ctacacccca atgccccaag agtacatcta 25320
aagacagatt ttgcttgttc gaagtgggga gggggaaatg catatggtaa gaagatgaac 25380
ttccccggtt gtgtgtacga ggagaaggct tccttaaagt tagctggcaa acatgcaagt 25440
aggagaaggc tgatatttcc tgtcaccaac atcttctgtc cagaaacaca gcctaaccaa 25500
taagcccagc tttcactatc tccagtcctg aaggacacct cgccctcact gcctgcccag 25560
ggctcagtga ttctcttaag ataaaggaag agtgggtgga ataatcttga cctcaagatg 25620
gaaatgattg cacaggaaga gcaaggggtg gggagattaa tgtggcctgc gaaaaaaaaa 25680

acggagaaag caagctgagg gatccatgtg ggcttttgtg acacccggga ccactgcaca 25740
ctatagtctt gctatagatg ctatatagtc ctgctatata gatgtggcta ccatctgaga 25800
ggtgctgagt caacatgtgc tgtctctgct ttttttttc tttaaggcaa gtagaaagca 25860
ggaatcaaaa ataccaaaca caagtaaaat cccctcacct ggaggccagt ttctcttcct 25920
gggaaagtat cagtgtgaat acacttctct caaatgcagc tgtagaatct gtaatgtctt 25980
tggtacctta aaaatggcgt tgaagcatgg tgtgggctct cagaggagtt ctttgaaggt 26040
tcatggagga gagaactgca aaaaacttgg caacaggggc tggggcagtg gtgtggtccc 26100
cagaaagcac ataatggtgc ttggttggta tgaatatgat taaattccca gcactgggtg 26160
gaagagggac aggaggagcc ctggagctta cccagcctaa caaaatattc tttaaaaaaa 26220
ttcactctga caagaggtct ggctgcttga ctcggggata aaggcacctt ccacactatc 26280
cttaccacct gagttttgat ctctaggatc cacataaagg tggaaagaga gaaccactcc 26340
acaaagctat tctctaacct acatacatcc tacacacaca cacacacaca cacacacaca 26400
cacacacaca cacacgcttt ctctagtaag aaagagacct acatgaagga atattcattt 26460
ttattgtaca tttggtcaat tcacctcgcc taaaagctca tcaagtgagc ctccccacg 26520
atatctctta gctcccacca agtcatcttt gtctcagtat ctgtaagttt cctcaagcaa 26580
tggataaaat ttgcatcctc agtattttgg tttgcatttc ctttattaaa aataacaata 26640
atccctttc acatatgttt tcatcatttg cattcctttt cttataattt gcttatttat 26700
ttatttttcg tttgttcatt tgtttgagat agagtctcat ttgtctcagg ttggctctga 26760
tctttctgct tctgcctcca gtgctggaat aacaggcgtg tttctcccat gctcagattt 26820
aatattttta tttaaaaaaa aaatcttttg gggcgggaga aatgtctcag cagctagaag 26880
tacttgttgc tctttcagag gatgtatgtt tgattcccag caccctgtg gcagcttata 26940
atcatccatt tcaggggacc cactaggcac atatgtagtg cacatacatt tgtgcaaata 27000
ggacatccaa acacataaaa taaacaaatc taaaaaacct ttctaaataa aaatatgttc 27060
ttgccaggta atggtggtac atgcctttca tcccagcact caggaaggaa aggcaggtag 27120
aactctgagt tcaaagccag cctggtctac agagtgagtt ccaggacagg caaggctaca 27180
caaagaaacc ctacctctga cctttgtcct cccaatatac atatatttag taaactaatt 27240
tttatttgat tagtttactt tttcttttaa agggttttcc gtgtgtgtgc gggagcgtgt 27300
gcgtgtgggt gtgttgttat tgttgttata tatgtgtgtg gtatgtgtan nnnnnnnnnn 27360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 27420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnaccacc accaccacca 27480
ccacagccac cacagccacc accctgctga caaggctcgg gtcacatgca cacaccacta 27540
tgccctccct gctctctatg tagttctggg gatctgaact cagaaaattt ataaactaaa 27600
tcatctctct agcctgtctt tttttccaga ttatattatt atttcacaag tctatcctga 27660
cttgtgaaat tttttgcttt gttttgtttt gtttttcgag acagggtttc tctgtatagc 27720
cctggctgtc ctggacctca ctttgtagac caggctgacc tcgaactcag aaatctgcct 27780
gcctctgcct tccgagtgct gggattaaag gcgtgcacca ccacgcccgg ccctgaaatt 27840
tttaaacaga ggctgggata tagctcagtt ggtacttgcc taatgcttgc tagaacacag 27900
ggagccggcc ctgtgatcag tcctcagcat cataaaagca gcatggtagc acacacctgg 27960
aatcccagta cttgggaggc agggctggga cacacagaag atcagggtca ttttcagcta 28020
aatagggaat ttgcagccag cctgtgcaac acgggcttct gttataaaac aacaacaaaa 28080
tcattttagc ttatcagttt ttctcctgtg gattctagat cttgtgactc accaacctta 28140
attattgggg ggaaaatagt atttttatt ttctaacact tttatcatgt ttctatgtca 28200
agtgctttat tttgacattt gctttaaaat ggggactaca actcttcttc ccggtgactc 28260
agcagcctcc gatagtaaac agtcccccct ttcctgccga ctcacggttc ctcagcccat 28320
atgctaacgt cccacttcct ctcagatgct ggatgtagac tcctcctcct gttccactta 28380
ttgattagtc tctacttgca ataatgcagc gatgcacatt ggtagcagtg acattcttat 28440
ggcctgtttc ggatttctgc tccagaagga gactataaat caagaccagg gcagcaaagg 28500
ggaagaggaa gtgggtaaat aagatggggg tcagcatagg tcctgtatga aggttccagc 28560
cagtccactg gggacttctt cttaggtaaa gctgaatgct aggtgagtgg tcagaaccag 28620
acaccaggaa ctctcagggg ataacctgtg catcccacaa tatgctcctt aaaggagaaa 28680
gacatctgac tcctaggacc attcctaccc cagctggttc ttcctctcac aaggagagtt 28740
ctagccacca ttctgcatgt ccaccctaac ctatgatgta ctctcccttt cccaaaggac 28800
cttaggagcc cagtgctgca gagttcactt cctctcccat tcagcactct cctaagagct 28860
cagcttgacc caccctacct ctgggaaggg aaggcccttt tatttacact gcgcttactc 28920
ttatcttcct tgtgtgcact ttttcctacc tcgatgcaga gccccagctt ggttttctca 28980
taaacttccc attgtctggt ctagcatatt cagatcaccc agggtctgct tgtgcttctg 29040
tcacccagaa gggtggagat gcctcctagc tgactggcaa ccctgtaagc ccacgtctgg 29100

tgtgaacagc tgatggcatg gagagacgta tgtctgttct gacttgactg ggatctgact 29160
gagctggtct tttttctggg tggtcctcaa cagctgatga atgaatctca caagatgctt 29220
tgcccaaatc tacatagatg agcttgatgg tggcaggcag gtggctgctg tgatgttgta 29280
gccatcaagg cttagcactg ttggttatct gacacatggt tttctttaag tatggaccag 29340
tgagaaacag ttcttaaaat taacagattt tttttttcca cataattatt tgaagttctt 29400
agctcctttg cagggagagg gggggggggc tccaatgagc ttgggcttgt tccctccttg 29460
cctaccatct caccccagga ctgttctgct tctgatacag aaaaggatct tcttatgggg 29520
gaagggggag ggacttgtgc ggtcccatca ccccctgaag atctgtcggt ggttaacaat 29580
accaggggag gcagtcagtt tctccagtgg tgtagctact ggccaggcat ccatgctcct 29640
gtcaataacc tctcatctat gctctcttat gaacaaccca aactaaactc atcggcatat 29700
gtgtgtgtgc atgcatgcat gcgcttgctc tcatgtgcac acgcgtgcgc gaacccacac 29760
tgcataaatg agagataaaa gaggggatta gaaacagaaa gagatcagca ggtatgtgag 29820
ggggacaaca gagggtcata ggaggtaaat gtgaacaaaa tgtgtgtgtg tgtgtgtgtg 29880
tgtgtgaaaa cctcacaata aagctgtttt ttctgtataa ttaatatata taaatctaaa 29940
agatttaata aagagcctaa tccttaaaac catgtttcac tgactggatt ggtccccatt 30000
gcacacccac tgtcccctag accgctacat gctttgcatt ttctcataga ccctgctttg 30060
gcagattgcc ccgacctgaa accttaccca agacgcacat aaccccacca aactgtgaac 30120
tgcagagtaa gccttctcac ccattgtgaa agtcacaact tccccgatct gcccttccct 30180
ttaggcctcc cctcccccaa ggcccagttc tcttgctcag aggtctcttc cttccctggg 30240
atggagcttg ggcgtcacag cagcaggtgt aactgcatct gtttgcctgc cttggatctc 30300
aggctagact gtgtcttgct cagttccttg gagcagcaag aggcgtcaga ggtaactttc 30360
ggttcctcta aaggctggtg tttatcaaag tctacctaac tgaacttctg gcctccagag 30420
gtctggggcg tgccatgctg aggagctgat ggacccatgc tatcatgaca gcttgccaag 30480
cagagaacac tttgttcttg tccctttccg gctcttctca gcctcagaca atcccaccac 30540
ccaccacggc aagctggaaa cttcctgttt tggtctgggc ctttgccatc tttaagttgg 30600
aaaatctttg tccttgatga atgctccttg ccctctgcat tgccccagaa gttattaacc 30660
tttttgtgac cactgtaatt tcttcaggat tcaatgttgc cttctaatcc tggggtcatg 30720
cagaaggaaa accaggtata gacaggaagc ccagtccgta ggagatgatg ctgttagaat 30780
ggctgatgtc actccctatg acacatgagg tgacaaacat ctgagagcta ttctcctata 30840
aggaaaacaa tgcccccaat cttctgccag tgtctgtggt gtcaatattc ccaataagac 30900
gcccactaga atggctgctt gggattgaag tcagtggaac catgggtaat aaccctggac 30960
agtgtgatat aaactcaaat tggaagagct gacaaagtcc tgactctaga ctttgagcat 31020
cttgatcaga gcctcatttt cctagttggg gtttctgttg tcatggagaa acaccattac 31080
caaaagcaag ttgaggagga aaggacttat ttggcttaca cttccacatc acagttcctc 31140
attgaaggaa gccaggacag gaacttcaac aggacaggag cctggagaca ggagctgatg 31200
cagaggccac ggaggtgctg cttactggct tgtttctcat ggcttgctca gtctgctttt 31260
cttatagaac ccaggaccac cagcccaggg atgaccccac ccacaatggg cttggctctt 31320
cccccatcaa ccactaatcg agaaaatgcc ttataggatt gcttacagcc tgattttatg 31380
gagtcatgtt gtagacaaca tttaatctca gtctgggttt ctaccctgtc tttgatcatt 31440
caattcccag ataaaagaca cacaacatga ttatttataa taagctttta atgcactaga 31500
gctgggtaga tatctaccct ctaaaccatc tgaatctact tccctaccca taaccccgag 31560
ttgtcacttg ccatgttcca tctgggccac tcttaactcc aagtggccag catggccatg 31620
tttctataat tcacctaccc catggtaact tctccacatt ccatcttctc tctttcctct 31680
cgtggttttc ctccaagcct gggaactccc aatccctgcc tgtctcaatt ctgcccagct 31740
atagactgta ggcatcttta ttcaccaatc agggataact tgggggagga gacaaggtta 31800
cagagctctt gggtctatgt gcagattctc ttgtccctgg gggcaaccag gccttgggga 31860
ccagtattta gcattattat acctagcaaa agaccaaacc tccacagagg cattttctca 31920
attgaagttc cctcctctcc aatgactcta atcttgtgcc aagttaactt aaaaatagcc 31980
cacatacccc tgtacctcag aggggtcagg caagttcaca gtcacttcca cctgacttcc 32040
ggtcttttca ggccctaaac ccttactcag agtcatttta attcagtatt ttcttttctt 32100
tttatcacct ttggggaggg attttttttt ttcccaagac ctcttttaca gacttgactg 32160
tgtgatcaaa cttcccattt ctctgagctt cggtttatgc ctctgaaaag tggaaggcaa 32220
ctgcccttag cagatgactt accccgggct tccagcagca actaactctg ctctcggcca 32280
agcagatcgt ataatcctct tccacagaca aggagagtag tgtagcctga gagcgttaag 32340
tatcacagcg ttcacactcc aacagcaaaa tgccccaatt agaagcaact taagggaagt 32400
aaggtttact tggggtcatg gttgagagga tatggtccat catggtgggc aagcaggata 32460
gtgggaccac aaggcttgcc ttactgtatc cacagtcagg aagctgaagg agatggacac 32520

ctgtgttcag ctgcctttct cctctttcat ttttttttt tttatagagt ctagacctcg 32580
ccccatgaga aatgctgcta cctgcattca gatggtaaat ctccctcagt taatcttctc 32640
tggaaattcc ttgacaaacc tgttgagaag tttgtctagt aggggattct aagtcctgtt 32700
aatgtggcaa tgaagattaa ctacaaaaca gtagagcctc gaattaactt ctcttgacat 32760
ttgacatttg taaattctag caatgctagg tctatgttcc ccacaggcta taattagaac 32820
taaatggaaa caacttcaac agtagcaact cctcctcctt ctccttctcc ttctccttct 32880
ccttctcctt ctccttctcc ttctccttct ccttctcctt ctccttctcc ttctccttct 32940
tcttcttctt cttcttcttc ttcttcttct tcttcttctt cttcttcttc ttcttcttct 33000
tctcctcctc ctcctcctcc tcctcctcct cctcccgcnn nnnnnnnnnn nnnnnnnnnn 33060
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 33120
nnnnnnnnnn nnnnnnnnac tttaaatttg taactactta tttgacaggg tctcattctg 33180
taaccagccc taaatttttt acaactttct acaaatggtg cttttggcct gggtataagc 33240
actcaaacca tgaatgggaa cattgtatat ttaaactgta gcaaccctct gaggctttgg 33300
ttttcaagtt gtcagtttga gagccctgac atgggcaagc taacagaata ggactgtgcc 33360
cagaaaaatg agtgctagcc cactgcccct accttcatct ccattgggaa ggaagatctg 33420
gggtctccac acaggctgct ccaaaacccc acccaaacta ccttctcagc ccaacaacct 33480
tgaattttcc atcaccttct tcacctccat ccactcacac agaacgaccg gacactctgt 33540
gactgtgttt atgttctcag tgtatccaaa ttattacagc tgagcactgc agatgaccca 33600
aggcagctgc aggttttata cagaaacctc taaaagtctg taaacctggg atggttttat 33660
atgatcttta taaatttgac acttttaaag aagatctctt tgaaaaacct ttccctccat 33720
gtacatgtgt atgtgcatgt gtgcacacac ttgcacgttt acatgctaaa aaaatgatct 33780
gcattacata aactgattct aataatgctt ttcataaccc atgttgaaat aatttttttg 33840
catgtaaaca actatgatat aggccatttc cagcctggtc ttttttttta aataaataaa 33900
taatttattt tgaaagtaaa ttgacttagg aaaaatttta aaatagtaca aagaattgca 33960
tttttctctc tcatcttccc actgagcgtt ttgacgtata atacactgat cattctctct 34020
catattgcac tgctgttttt ctgaattatt ggagaggaag ggtctgacag taccctatta 34080
tcactatctc attcatagcc atattcagcc atcaaaatta ggacattaat gctggcctac 34140
cgtccactcc catagctcca ctcaaatgcc tccagcagtc ttagcagcaa tataaaacat 34200
cctatttgtg taactatgtg aactataatt tatacacata aaataacttt cctcctggac 34260
caatgtttag tcgccaaacc tgggggtgct gttgtgttgt gccttccttc atgatattga 34320
ctttttttt tgtttttgag aacttggtag taattttgta ggctgcttct tggtttgagt 34380
tcttcagggc tttcctcgtg gttgggtcca gttctgcatt tccatttcca gcaggacttc 34440
tcagtggtgg tgctaaactt actactcctg ctcaaggtga catgccggct gctctgcgca 34500
gttgttagag tacgctctga gtgtgaagct aagataacat ctgaaggggt ttctctttgc 34560
aaaattagat atttgttatt tggggtctta atattccctc cctctctccc tccgtccctc 34620
cgtccctctc tccctctgct taggattgaa accagagtct cacactattc tgttcttggc 34680
cactgatcta tcttcacacc ctcaggaatt cagggtcgtc ccaggcatca tagtaagtcc 34740
aaggccagct tagaccacac gagtccctgt caaaaagaaa ataaatgact gcagtcaagg 34800
atagcagcac agcttgaaat cccagcactc agaagcacag gcaagaggat caggggtttg 34860
agtccagctt cggctgcata ataaaaccct gtctcaaaat aaacaaataa aagactgatg 34920
actaggctga tagctataaa caacttgaat gtcaactgcc tagcataccc aaggctctga 34980
attcaacccc taacaaagca aataaaataa aactagaaaa ttatacacac acacctgttt 35040
tgtggaaata atccgagact atctaaatgt gtcatctctc aaactctcac ccacctggtc 35100
ttagcatccc ctggtgaagt gcgtctgcat cagtagttac catgatagct gccaaatgct 35160
gactgcctct tctatcattg gtcccacagt tattaattga catttcgctg tagggaaagc 35220
ttcccttctg atttatttat gtaaggataa agttgtgaat tctcatttta cccagtggat 35280
tgtaactatc ttcattattc atttcaatgt tcagagtgta ctggctgctc ctcctgctac 35340
acttggtggt tcctggctct ttttgacaaa ttgtcttccc gttcaaagca tggcttttct 35400
ccctagaacc tactagtgca gtggcccttt aatatagttc ttcatgctgt ggtgaccccc 35460
aatcataaaa ctaatttcat tgctacttca ttaactataa ttttgctact gttatacatc 35520
agaaggtaaa tatttttgaa gatagaggtt tgccaaaggg gtcacaacca atagtttgag 35580
aacctctact ctagaaatat atatctcagg ttctgctgtg ccttcccagc cccagattgt 35640
aggctcttag ctgaatttaa gaacctctgg gtctttgtaa tggggaaagc acccaagatg 35700
tgaacagtgc tcattgctgt tagtctctgc cctctcagca gacagaacca agaaggtggg 35760
ggaggaagtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tatacaattg 35820
atactgttac ttaccattca aatgccaaac aacagaattg ctctgcaact ccctccctag 35880
cctggcttgc atgcttatgc ttgttcctta gcaggtnnnn nnnnnnnnnn nnnnnnnnnn 35940

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 36000
nnnnnnnnnn nnnnnnacca tgaccccagc tcggaagttg gctctagcta cttcaaaggc 36060
gacgaaacaa tgaaactgtg ctaataacca gaggaaaggg agctgggcat ggtgggtcat 36120
gcctataatc ccagtacttg ggaggctgag gcaggaggat ctcaagttcc actctagcct 36180
gagctatatg gttccagagc aaaatccttt ctcagaggaa gaaggaaggg ggtgagctgt 36240
cttctttggg gatatccttt agagaaacca ctctcagccc tagaactaaa gacaggtttc 36300
tgagtctaaa gcagtccccc aaactattct cagcccaagc tgccccttc cccaggcttc 36360
tgggtctctc tagctggcca cgcctcctca gtctattagt gttctagctg taccagatcc 36420
ctgtccctgc ccctttcccc tttcttccac cctacctccc tcattcaagt acagactctt 36480
accctgtact gctgccaaag tgatgtcatc ctgaccagca gtggtttgca cctgtaatcg 36540
ccacactcag atgatgagtt tgagtcaagt ctgggctata atgtaagaac ttgaatcaat 36600
caataaataa aagttaaaat atcaatccta catgaatgca caccccagct cttgagtgct 36660
cccaaaagct ttcaaaatta ttgttaaaaa taatagtaag ttctacaaaa tcccagaggc 36720
cccagaaatc atgcgctctc cccttcaccc acatgaagtc tttttggcca aggcagagga 36780
ggtgagtaga cattgagaag aaggcagacg gaaggtagag gaaacttcct gatgactgca 36840
gagataggag agggcagagg ggcctgtgga gagagactca ggagaagctg ggtcctagac 36900
acactgataa ctgcataggg agatggcaga ccatgctatc taaaacacaa aggcctggtc 36960
aggccagctg aactcctgag ctgcccctgc tgccaccccc cttccacagt ccccccttccc 37020
cttcccctgc cccctgccca gcacagctct cagatcttct gtgtaagttg tatctgtgga 37080
acacccccaa gcttggactg agtctagccc ctgcctgtcc ctcaatcacc gtcttttcaa 37140
ctggcagact cagggtcctc actaagaagc agaaataata aacacgatca ctggcacccc 37200
tctgcccgca gacaaatgtg ttttcaccct gctttcaccc tgctgcctaa gccttgtctt 37260
ctgactttct ccttggacct gaagagggtg ctttgatttg gtgtgaatct gcatggagct 37320
ctttggtttg ggttccccac ttcccacttc ccaccatcta taaaatctcc atggtaactg 37380
aatgggccgg cctcagaatg gctaagtagc taaagatcgc cttgaatccc tgatcctcgg 37440
tctctgcctc ccaagggctg gcatcttggg cctgcagtcc tgctctggga agtcttttca 37500
acctaaaagg aaggctaccc caaccttcag gagtggccac ccggagcacc acacttctgc 37560
tcagcggctg gcagacttcc tcccctgtgc ccagtctatt tcagttctgc atttagagat 37620
gtcagacagt gctgacatta aaactatccc ctgggatgtt tcagcttgtc tagggctctg 37680
accgcccagg cctgctcttt cctgctcccc acccccaccc tgcccccagg gtattttcag 37740
tcatccccag gactgcatct gaaaaagcag ccagcagtct gaccaccagc tgaccttcag 37800
atgaccttag tcttaaaggt accagtgccc tccctcctct tcctcaaacc tggacccccc 37860
cagcctgaag gccaaccaac cctccatacc caaaaacctt tgtgcatggc tttgtacctt 37920
ctgggatatt gctccctggt agatgcagat gccaagagat agaacaagat tttccctata 37980
agatctagct gggcagtcca ctcccctcca cctccctgca ccttttatnn nnnnnnnnnn 38040
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 38100
nnnnnnnnnn nnnnnnnnnn nnnnnnnngt tttatggagc agatgtatct cagaacagtg 38160
ggctgtggtt atgtaccgta ccagtttccc ctgtcaacag acctggctat aagacagggt 38220
ttttcaaggc tcttgtgaag ggggggggtt ggtaaagtag gaaccacagt tatatcccaa 38280
ggcagagtga gctgtgcatt caaagcttga ctccttgggt tgttgtctat aaaggcctta 38340
taaaaaacat tgcttctagt tactctgtgt agacatatta actaatcaga gacaaggact 38400
atatttcatt tcgtgaacag aactaatcct tactttattg tcttgtgtgt gacttttgat 38460
tctgaacata ataaattcat gtccagtctg ttgattaagc aaatcccatg gagaggtcta 38520
tcagcctcaa gcctgggata tcccagttac aagcattatt ttaaagtagg aagagcctga 38580
ggctgagggt ggagcacagt gaaagggctc gggccacacc atgcaccaca cccgctcatc 38640
ttcaggttct ctctctcttg gatcaaactg cctgccacat gctatagtta atctgttaag 38700
gagagaagga attcatgcct gccaagtgga agaggagagt caccggggag caggggcctc 38760
acggtccatg ggtattgtgg ggtatcattc ctaaggggca cagtgagggt acatcagtaa 38820
acagctcaga aggctgttag aggcccacag ctaactcgag acctttgggg ttccttctct 38880
atttcaaaat gacttggtgt atcttttaag gaaggcctta gctgcagaat gaggaacaaa 38940
agctggcggc ctagacgcta agaatccaga gccacgcccc tctgctggca gggaactatg 39000
aactaggcct tgatctgagc ctcatgttcc atgtgacaca caggtgcttc tctgccatca 39060
tgattgtcct ctggttcatt cactcctttg agtttagcca gtgagagccc cagccaggtg 39120
gtaagacagc gagaggaaga tagtgaagaa gcgggggtgg gtcctctggc actggctccc 39180
tcctgctggg agttctcagt gggctccagc actatccttc tgctgggagg ttccaggggg 39240
gtaggggtaa actgcttccc tccgctgtta gctctagaac tcttcctctc tacctaccaa 39300
actctcccca aatgccagtt ttatatggcc atctgcttcc tgctaggact atgactaatg 39360

cctatgacac ctttcttgac cacattttac attttaaaat agtcgaaaaa gcaccacctt 39420
tgggattgtt gggagaaaag ggggtacacc aagccccacc tctcactgtt tctacctcca 39480
agcccaagca cacagaaggt atctaaccta gttttatttc ctgttgctgt aataacaaaa 39540
agcaactaag ggggtggcgg gaggaaaggt ttgtttttgt tttcactcat aattccacat 39600
tacggccctt tgtaacgggg aggtgaaagt ggtagccagg cagaggtggc gggtgctgtt 39660
ggtcccagtg cttgagagag acagaggcag gtaaatctct gagttcaagg ttagcatggt 39720
ctacaaagtg agttcaagga cagccagggc tacacagaga aaccctatcg taaaaaaacc 39780
aaaacaaaac aaaaaggcaa tactcctctc ctattccctg ccccccccaca cacacacacg 39840
ccataggcca acctgatcta gacaattcct cactgagact ctcttcccag ataattctag 39900
actgtgtcaa gttcacaatt aaaactacaa attcttgcct gtcattctga ttccagtgtg 39960
gtaaccccac agaatcaggg gttgccaggc tcctcagaat tacaggaagt acggtagacc 40020
gagatgtagt atgtagaccc cgctcctgtg gggaggggggg cctcttaggc aggtgaatat 40080
gtgcaacatc tcatgtggaa ggggacctct aggctcatta ctaaaggaca ttattgttcc 40140
tcctggccca cagtcagtac actgccacct agaagatctt cccacgtgtt ggtttaggct 40200
tagtggggtt tattggctac acagagccgg ccattcacta tccactctgc ctggacatag 40260
ggccttttaa gcccatctta tgaagacagc attttatact gtgcctccaa gagacccctt 40320
tctctggcct gttgggaccc atatctccac catctcctct gcctctgtct gctgcctgtg 40380
aagacaggct gcttcctgct ggcccacatg gcataacaat tgaaatttat ctcatgggaa 40440
tttgcttccc ttaggagggc ccttgtatcc tttctgggta ctagaagtac aaagcactca 40500
gagctgtccc tgagccccag ctcctgggta gtgaatctat attatagtgt ggccataagg 40560
cctttgcaag gacatttctc tgtcaaatgg acccagcacc cacctggcat ctctgcctcc 40620
tgtctcatcc ccatctgtct catcagttgg ccttgtatat gtgcatattc attgcataca 40680
taaataaata agtggatctt gttgctctga ctgagtacca caaacagggt cacttttaaa 40740
gattaaagtc caggcatgta gcttgaaggt aaagcaaatg cctagcatga gcaaggtcct 40800
agccagacac ccgtgaatag aggcatacag cttatttggc tcacagttct ggatgtagga 40860
tcctaggagc aatgtgctga cttctatcta gggaggacct tcttgctgta tcataacatg 40920
gcatggggta tcatatggcc agacagcaag ggtgagctga gattgctttt cctctgcaac 40980
aagagccact gacaccatca cgtaaacaca cacacacaca cacacacaca cacacacaca 41040
cctattacaa aggctttacc tccaaacacc atgtacatat gaatacaggg attgcctttc 41100
taataagtgg gaagacagtc caaactatcc cacacttggg gagttgtgcc tatgggaggc 41160
agaggaaagc ctcccttctt agcagcgccg ctccaagctg gctcatgctc atgcatccac 41220
ctggaggagc ctgcttcccc atccttgaca tttgacctca accctacact ccccatcttc 41280
ccggaagcct gacccaccta ggtggtcagc catcctggcc tgccaggccc tgctccctgg 41340
cctccatgac aatgcaagca tgagcagggg ccagggaggc atggctgact aacagctctc 41400
agcagtcaca actgcctcca gctcctgcct gctccttttg gggatccctt tagcttctcc 41460
ttcctctgct tgcttaatac acagcctatt cacccaccat tgtttacccc atccccactc 41520
tgcctgccac ggccaccttg aacatgtcta caattcaata gtgcagcata gagcctcgtc 41580
tctgggtttg gttctggctt ttttttttc ctgctgtata agataatgtg catctcagcc 41640
tgcacgcccc ctaagcccca atcgcaccag tgttcacaca tgcagtcctt aggcactctc 41700
cactccctga tgagactcct tccccagatt cactcaggag tcagcccatc tcaaatttct 41760
ttgatatttt cattctcttt ccgaatgccc agtggcattc gaatcgattc gaatcatcga 41820
ttctctctaa ccagccctga caagaccaca cacactttac ctgaagtgag ttcacaggtc 41880
ttacatctta tcaaattatt cttgttattc cattcttgtt tgcatttgaa ccactcactg 41940
gagaatattt tggaaatctt ggcaggggct gccactcagg tggtttaaca gtgctcacaa 42000
gagattgagc agcgtggggt tttgaggaga caggggagtt tattgaaaag aataaacaga 42060
tgggtttgct aagctctcta taaatttact aagacaacct aagccataca gcatgagtgt 42120
gtatatatgc tctctcacgc tctctctctc tctttgtcac acacacacac acacacatac 42180
acacacacac gcacacgcac ggatttttgt gtgtaaaaca catgagcaca caagtacact 42240
tggagagcgg gaagcacgca gagggaggcc tggagcccat atggacagcc cgagtccaca 42300
gagactgttc tcaccaaagg ccgggtgact cactacagtg tagaagaccc agaacccaag 42360
gttcagagac aagtctgaga gcttcctagt tgtagcagct atttgtctca ttggtgtaac 42420
aaaatacctt acaaaagcaa cttgaggaag aaacgagtta atttggctta cagttcaggg 42480
gatacagtct acagcaatgg ggatggcaca ggacgggaag cagaggccac tggccacacc 42540
atgtccagag tcaggaaacc aggacgaacg ctggagtcag tttgttttct cctctttact 42600
cagtcctgga cctcagacca tgggatgatg tcacctacat ttagggtggg ttttcccacc 42660
tcagtaagcc ccacctagag aagcccaaag acttgtatct atggtgagcc ttaattccat 42720
caaattggca agattcctga gtctgtccca ttcacctgag agcataccaa ccctcccaaa 42780

tcaagcagtc aaaccatcag gaataggaat ctaatataga tgtgactgag tcaaccacct 42840
gccttccctc ctgtgaggac tttctaactc aaacccatgc tcctattggt ctggtttacc 42900
ccccaagcaa tgggtttctt tcagcaagaa ggtagctagc tggccctaaa gaaatggttc 42960
agcctctgat ctctaaacct gaagatggac acaagtcatg ttaaaatgat tggggcagct 43020
ggatgtggca gctcatgcct gtaatcccaa cacttaggga gctgagacag gaggcttgcc 43080
atgagtttga ggccaacctg agctactatg tagaacattg tcttaaaata aaataagtta 43140
aataatagtt agacaagagt gaactatggg cagacacttg gcccatgtgg agggatcatg 43200
accccacaaa gccgcaggag actcactaga gagagtatgt atctcattag ctgcagcaat 43260
ttggtctctt gctctggtct ggctgtctgt agggatagcc atcttagcaa gggtatatct 43320
aaatggatga aggggcatct tggtggggtc tgggctaact attcagggaa gaccttgaat 43380
ttcctactgg agctgcctct tccaccaaca tacaaaagag tcctggtccc agagtgtgag 43440
agcagtagag cagagtcacc tccatcatca aaagcagagg cacaaactgg tgatgatccc 43500
tgtgttgctt ggacaacttt ctgacttttg gctgaaaaaa aaaaaaaaaa caaacaaaca 43560
agaaacatag ttaagggctg gatgggggac caagtggtta agagccttga ctgctcttgc 43620
agaggatcct gggttcagtt cccagcccca catggtggtt cacaacgatc tctaactcca 43680
gtctaaagga tatgatgccc tcttgtggcc tccatgggca ctgcagtcac atagtacaca 43740
aaacaatttt atattttaaa aaagtaacta agtctagcgt ggttactagt gcacaccttc 43800
aatcccagga ctcaagagac aggaacagaa aaatctctgt gagttccagg ttagccaggg 43860
ttccatagtg agacccactc aaaaaaacaa acaaaaacat agaatttaaa caaatggaaa 43920
aaaaagtaca agcatgaatt ttcctatgtt taaaagcaat aatcaagtgt ggtggcacaa 43980
gcttgtaatc tttaagccct tgaggatcta atgcaggagg gtggttgcaa gcttgaggct 44040
acgttgaatt atatagccag accccatttc aaaaagcctt aactaaatac agtaaacaag 44100
ccacgtgtgg tagcacgtgt tcttcctcgg agggttgagc caggaggatg acatattcaa 44160
gcttggtctg gatagtttag aggaaacctg cttcaaaatc aagtttaaaa cggcagactg 44220
ttgatgcctg gtgacacaca tctgtcatcc cagcacccag gaggcagagg caggagatgt 44280
aggacggatg tttccttagc ttgtgtacag tcctcgcttc aatccccggt gctgtaaaac 44340
ataagtcatt aataaacaaa tagtcggggc ccggggtgtc actcacggta gggctatgtg 44400
tgctttgcat gtgtgagaat ctgagctcta ccttcaatac caaaaataaa atataatttt 44460
tagaaatggt cttggtgggg gggggagcaa aaatagctct gttcaccaca cacacaccta 44520
aacttgtaaa aatagaactc ttggcccttt gagattatag gtgatttatt ttttttaaaa 44580
aactggtatt ttctaatttg gatataataa aaaaatatat actaatttgt atagtaaagt 44640
acttatagaa atcattttta acagatggag agaaaaaaat agttttaaaa atacttctat 44700
ttgcaccttt gcaaagggct cctgggagcc actggctgcg ggatgtgagg cgtgagtgga 44760
gagcttttct gggatatggt cactcctagt ctcttcagtc tctgctccag gtggcctttg 44820
tgcttgggtg acattctagc taaagtgtgc tctgtccggg gcttctgtcc tcaggcatcc 44880
gaaggtgtcc ccatgaggtt cttcacgaag ctggaccagc tcatcgactt ttacaagaag 44940
gaaaacatgg ggctggtgac ccacctgcag taccccgtgc ccctggagga ggaggatgct 45000
attgatgagg ctgaggagga cactggtagg aagggaagga agggatggca agtgtgggga 45060
ggtgacaaaa ggactctgtg gtccagcctc acgggtgctt ctagactaga gtacctcttt 45120
ggaggttgaa ggatgctttc acaggggtcc cgtaccagat atctgcattg aaattcataa 45180
cagtagcaaa attacagtta caaagtagca acaaaataat cttatggttg cggttactgc 45240
aacctgagga ccgtattaaa gggttgcagc actggattag acagtcttgc tggcacagaa 45300
atgattggcc gggaactcat atgactctga agtcaggtgt agactagaat tcctgctccc 45360
cggcttgttg cctttgtggc cttagggact tctgactctc tgaaccacaa tcttgccttt 45420
aaagagatta gataagagta gcccatatta atctctcagg atgtcacagg accagatgag 45480
cgcgtgaatg gaacaccatc agcagtgcac tccagaaatg aagttgccat cagtttcttc 45540
actatagtga aagttgagag gctccatgca cataagtgag agggtgggga agtgcttctt 45600
catgtaccat cactaaaggc aaccaaaatg gtcatggtgg gacatgccac aggctattcc 45660
aaaatttagg atgccaaggc aggaaaagca tgagtttgag ctagcctggg ctccatagtg 45720
agacaccacc atctgtaaag aggcttctaa aatatgtcgg agaaaaagag tttccaagag 45780
tttgtgaagg ccagcacaga gacaaggagc cagtcttgga gagacaggta tgcagaaagc 45840
cctgtagcag gacacccttc ctttcatctc aaaggttcct tgcctgacag aagcccacat 45900
gttacctcgt gagacgctca gcaggtcagt gagttcccgg caggttcccc ggggaaagca 45960
ttagacatgt agtgttgaca tcccaggcct ggctcctacc tgtccctgga gagaatatgc 46020
caggtgtctt gttatactgt gggctagctg tggattgcct ggcctctgtt tcaaccttct 46080
ccttcctgaa gaccctccag agggcagcca ggagagctgc ctcaccatcc tgagggagcc 46140
acagaggcac atggtgaagc atgtgtctta gatggccatg gtttgaagcc tgcttagatg 46200

ggggactgcc ctgtctcctc tgcctgacct ctagagatca gcagctgagg tatagctgag 46260
accaaaattc acaagaaggc cacaaggcca catggcagcc ttgacctggt tctgtagcta 46320
tgtccttgac agggcaccac tctctcttcc ctcaggctac catgcttttt ctcagcttgc 46380
tctgtcactg gcatcccatt agaggcaaac ccttcaatgt gtcctataaa ggctctggag 46440
agaatctgtt tcctgacctt gacttcacct ccccaaagct cctggtcacc tgcaagaccc 46500
aaagccaagg gacacccagt gcctggacca aggcatctgg tccatcttgg gtcttctcga 46560
gggcctgcag ttccagggct caaaataggt tcagatctgt ctcccacagc cagagggaaa 46620
ggaatgtaca aatacactga agaaaaaaac ccttgtttct gtcacctgtg tagtgctatg 46680
gctgaccaga gtggccatca gggggcaggg cagggcttgg gcaggaagtc ctgagtctaa 46740
atgtcccagt caccttgact ttgtgcacat gaaggacatt tgaaagggtg ggcaagggtt 46800
acaatctccc agatgaaagt caagttgttt gatattttag gtctcagaag ccctttgctt 46860
tctaaatttg tttctgaggg tctagaagag ggaaaggaag cctgaggctg gagagactga 46920
agagaaggaa ggcggaagtg agagacagac agacagacag acagacattg ggtgggcttg 46980
attttctgcc cctagtccct gactatactt tctagacccc gagagccaag tcatcttgct 47040
gatctgccaa tcaatctgtc caaggccagc tatcttgaag gtgtccctgt gcttctcaga 47100
accaactgac cctaggaggt acttgctgag cagcatggta gccagaacaa caagaccaca 47160
gccaggggct tcagatctaa gggtctgaaa gcgaggtcac ctcaagaggc tccctatgta 47220
gtcagcatgg atttaggcag gaagaatggg aatagtgggg cagactggga cctggaactg 47280
gggagcctgg acctctccgg cttgtggtag aaatgcggaa cagtgctgta cggacagcat 47340
cctcactgat ggcatttgtc cctgctgttc ttttagtaga aagtgtcatg tcaccacctg 47400
agctgcctcc cagaaacatt cctatgtctg ccgggcccag cgaggccaag gaccttcctc 47460
ttgcaacaga gaacccccga gcccctgagg tcacccggct gagtctctcc gagacactgt 47520
ttcagcgtct acagagcatg gataccagtg ggtgagtttc tacttggagg gtcccctggg 47580
aagccggcct atgcctagga cccagcatct agcgtccggg catgctcagg cctggactgc 47640
ggggagcagc agacctgctt attcctagta aagttgaaag ccaatgtctg tctttctagt 47700
caggactatc tccctgcccc caagtcatgt ttgttcttag tgcatttcgc tattaagatc 47760
gtcacatttt tcttgtgaga ttttatgatc ttcttggagt atattttaca ttcataactt 47820
ttaggtcttt ctgccctgta agtacaagat tgctttttg tgtagaaggt cactgccgtt 47880
ctgtggtcag atggcacctc cacacccacg gtcgcccacg ttcctgaagt gtttctcatg 47940
ggaaatcact ctacagggct gctcctgcct gttgtactac atcccaccac aggcagctgc 48000
gggaggacct ctccatgcca cagccacacc ttagtctgag aaatggtagc cccaaacccc 48060
aggaaaagca atgtgactat ctacgtgctt gctcgctccc cacatgcctc ccatggggct 48120
ccaggatggc accaccaact tatagcttca ctaatggagt acacagtgct aagtgacatt 48180
cttagaactc catcaagtgc taagagaaag gctgactgct aagagccaga gcccagggag 48240
tacatggtct cctgggatat gtgacagtgg tgaggggatc aggtaacaca gcctaactgg 48300
aaggagacaa tttgggtccc ataccaggcc aaatccaggc ctgtggggac cactatctgg 48360
tcttgcattc ctggccctaa tcactcacag aagtcaggca tggtggtata tgtctgccat 48420
tccaacactt gggaagtcaa ctggatgttg tcttatacct atagtgccag cactgagaga 48480
aggcttgata ctgaaagatc aagagttcaa agccagcctg ggctacatag gcagatccta 48540
tcctcccacc ccacccccaa ccctagaaaa atataaaatt aatactacat tgacttccca 48600
atggaggcct gagtgactcc aacccctcta tactgtagtg ctccaggtca cctttgagag 48660
tttggacttg gcatgagggt cagacctatt tgttatctct aggctagctt gttcttccag 48720
ccctcggctt acagctgtgt gaactaagga gctagagaac tggccctcca cgtgtttctg 48780
aggctgtctc tttaaaacac aaaaagtcag atggaggctg tatgcacctt tgatcccagc 48840
actcaggagg cagaggcagg cagatttgtg taagtttgag gccagcttgg tctacagggt 48900
gagttctaga acagccaggg ctacacagag aaaccctgtc tccaaaacaa acaaataaac 48960
aaacaaacaa aaagaatgac catttcccag cccaattacc cctccaacag cacccaaaag 49020
tgttctcatc ctaagaggta ctgtactctg tggcctctac ctgactttga tcaattcctt 49080
gctctgcagt cccacctgag gggactttca ttatttcttt gtttctcagg cttcccgagg 49140
agcacctgaa agccatccag gattatctga gcactcagct cctcctggat tccgactttt 49200
tgaagacggg ctccagcaac ctccctcacc tgaagaagct gatgtcactg ctctgcaagg 49260
agctccatgg gtaacggaga gccctgagag aggggtgggg gagcttcatg ggtaatggga 49320
gcccctaccc acccaggagg atggccacag caagagaaag tgctcattag agtgaccctg 49380
ggtctcctct ctgtccagat gtctctgcag cactcacagt aattggccca ggtggagtct 49440
ggaatgttcc aggcttgttg gaagctcttg ctctcataga atctgagctc taactgagct 49500
gggaaagttg atcatttgtt tattcctttt agggtattgg gggggcacgg atgtatgcat 49560
gctggtatgt atgcatgctg gggatgtagg gcctcatgtg tgctaaatac atgtgcctat 49620

tctgtaatgt tttcttgttt gtttttaact caaattaatt agaggcagtt tctctgttta 49680
aaaaaaaaag gtggaagaca ctgccatctc atttgtgttg ggacttgaga tatatatata 49740
tatatatata tatatatata tatatatata atttatttat ttatttattt gtgtatgtgt 49800
gtataagaaa gcgggggaca ctcatgtgcc atggtgtgta tggggacaca cgtgtcctgt 49860
gggggaacat tgtgtcaggg gaacaagcat gtgctgaggg tgggggggata tatatgtgcc 49920
actgtggtga tttggttggg ggagacatgt atgtgctatg atacctgtgt gtaggacaga 49980
ggacagactc acacaggagc cttcttacct gctgagctgt ctcagcagtc caagccctcc 50040
aggctgtagc cactcttcct ccttgctaca gtcccacatc cagccaacac ataaaggctc 50100
tggcaggaaa taaattaaat ttgctttgtg tgtgggtgct gacggctcaa gtcttccggt 50160
gatggtaatg tttaaagcaa gccaacatca gttctccagt gccgaagtta ttgaatgact 50220
gaccaatggg taacacttag gatttttaaa aattattttg attgtacatt ttgaattaca 50280
ttaatcttaa aataaactac aaacataaac acactgtggt tggagctggg catagtggca 50340
catgccttta attccagaac ttgggaggca gaggctggca tatctctatg actttgaggc 50400
cagcctggtc tatataatga gttccaggac agagagatcc tgtctcacaa acagagaaaa 50460
ccccataact atacttttat tgtactggtt ggtctacagt gtaaagaatt ggcaaagaat 50520
atgaaagatt acactgggaa gaaactgaaa gccatccaga gtaatgaagc aaaccctca 50580
caaatgtggg aaacatttca catgggtgag ggcttgcagc cagcctgtgc taatttatgt 50640
tttggtacgt gagcacttag agcagtttcc agttctctgg tgctctacca atcttagctt 50700
ggttaatatt cagggatgag tcttccatca ccggtaatct aatttgccat tctatttaaa 50760
ggctcttaaa ggcacaggca gtgcattggg taaatgtggc aaataatttc tttgacaaat 50820
cgaccaattg tcagattggc ctgctagtca tttgtttcaa tgagaactgt tttttctcaa 50880
aggatgctct tgtacaccgg ctagaagcag ggctgtcatt tttataggtc tctgtggtat 50940
tttgttgttg ttgttctgtt gttgcaaatc atttatcact gagggaaaat acacacaaag 51000
ggccctttct ttaaaagtat acatgtatca ttttgtgaca gctcataaga agctgttttt 51060
ttctgcctgg acacaggtcc tgacctgtgc tgtgtccttg ctaagctttg tcagaccctt 51120
ccacagctcc ccccaacaac gagttcccca gtacctgcct cacctcatca ctatggtgac 51180
agcagcctct gatgcgcctg actctctggc acattatggc agtgttaaaa gcttccatct 51240
ctctctttgc tgaataatga acctcaggtt gttcaagaga ccggaatgtt cttcacctgc 51300
ctgcacacat ctcttcactt tcttttatag atcaggtagg gactgggcgt gtagatggaa 51360
caaactgttt tccgttcccc agccatctct gcaggtgcac tccacataaa tcaagtgtta 51420
aaagtgcttt gattaaacag gacaggcgcg ttcttgagtt catctgttca catactgtct 51480
ggcaagcgct gactgagggt ctcctctgta ccctgttctg agaactaaca aaagacgaat 51540
caacatacag aaaactgtta tttagtgact gattaaacta acgaaggcat gggctggaga 51600
aataactcag cagttaagaa catttgctga tcttgcagag gacctgggtt gggttcctag 51660
cacacacagg gacagttcca gtcccggtgt gtccttttct cacttctgtg gacacaagtt 51720
ttacacatag tgcacacaca tacactcaca tatataaaac agaacattta aaagtatgtt 51780
taaataacgg aatcatttat ataggttttc atttacatag gtaaataggc aaaaatctgc 51840
attttattgt ttctaagttt taatttattt ttctctgtgt gtacatacgc atgcctcctt 51900
atctgtatgt gcgtgcactg tgtgcatgca tgaacccaca gagaccagaa gagtaccaca 51960
gattctctgg agctggagtg attgataggc tgttgggagc cactccacat ggggattcag 52020
agttgaactt cgttctctgc aagaacagcc agctcttaac tgatggcttt tacctccagc 52080
caccttttcc tcatttttaa aatttccttc cttccttttt gagacagggt ctcaatactt 52140
agctcatccc aacttgaccc cactcttctc ttgccttagt caccacaatg tttagtttat 52200
aagcatgcgt cactatgccc ggctttaaat aaactcaccc ataatcccag cactgaagta 52260
gacaaaaggg aggatcgatg gggctgactg gccacaagcc gtgcttcaag ttcaatgaag 52320
accctgtctc aagggaataa ggcacagagg atagagccat acgcctgacc tcctcctctg 52380
gcctctaccc aggcacatgt gcatacacac accacacaca cacacacaca cacacacaca 52440
cacacacaca gagagagaga gagagagaga gagagagaga gagagagaga gagaaacttt 52500
ttcctctttt tttttaaaaa tattatttat ttcatgtata tgagtacact gttgctgtct 52560
tcagacaccc caaaagaggg catcagatcc cattacaaat ggttgtgagc caccatgtgg 52620
ttgctgggaa ttgaactcag gacctctgga aaagcagtca gtgctcttaa ccatctcttc 52680
agccccacaa agaaactttt aatgagcaaa taattgcttc caagtaaata ctactaatat 52740
atttctaacc atactataca aggaattatt aaagaacgga taataggaga ataaaaaatt 52800
ataagtcact ttataatgct atctaatcca tctagaacaa aaacactgta ataatgcaaa 52860
agagcgcagt gcctagatta aataaataaa atgcagacca ataagtaaac tttatagcag 52920
cacatggaaa tgacgaaatt cctaacaaaa agctcaagat gggcagttta tttaaagtga 52980
aatacaggag aaataaagca cagaaagata ctcaaaggca tagaagttaa catagggggg 53040

ctggcgagat ggctcaacag gtaagagcac ccgactgctc ttccaaaggt cctgagttca 53100
aatcccagca accacatggt ggctcacaac catccgtaat gagatttgac tccctcttct 53160
ggtgtgtctg aagacagcta cagtgtactt aacatataat aaataaataa aactttaaaa 53220
aaaaaattaa agaagttaac atagaagccc actcaggacc ccactcagtc ctagagtatg 53280
acattattat ggacattaaa aagagaaaat tcagcagtag tgtgcatgca ctgcatatat 53340
acacaaatcc ttgagtttca taccaaatgc ctttagacca cttgtggctc tgcaaacctg 53400
taatcctagc acttgtgaaa aggtcagctc aggaactttg ggaaggtcat gaaactcttg 53460
cccctccaga agggagaggc taattaacat ttctcagacc acagggcggg aaccgacctg 53520
cgggtgggga cagactgttg cccatttcca gactagggaa gtccttgtca cctcattccc 53580
taaagaccaa tcaatttaaa gggtgcactg ttccgccaat catattgtgc ctagttgctg 53640
atgctctatt ctgcccttag aaaccgtata aaaactagcg aaggggtacc aggggtaacc 53700
ccctctcctt caggtctggg acaatcccac tacactggaa caataaattc ctcttgcttt 53760
ttgcattgat cacagctcca cttcgtggta agctaagact ccctggagtc ttacattggc 53820
aaatgcaggc aaaagaatcc gaaactcaag gtcatctaaa actacatagc aagcatgctg 53880
ctagcctggg ctccatgaga ccctggggga ggggcagagg gagaccgttc agaagacagt 53940
caagatgttg cagcagcaca ggcagcctgg ccaccagtgc tgtcaccaga catgttaatg 54000
ttggaataaa gcctcaatca tgactctccc agttttataa ttggaaataa gaaaggaaag 54060
actataggaa caactgtgtt cagaacacta tttataatag caaagatctc agagtaaccc 54120
aaacttctag acattgattt gggaagatct cttggcagct tattttgaaa actttacaat 54180
gttaaatatg taaaaacaag gacagttttg ttttttgttt tgttttgttt tgttttaggg 54240
atatatattc atatatgtat atgaatgaaa acccaaactt aaaattcccc actatgcttt 54300
aaaggctttc tgacaataac agaaagagaa atagagaatc cataaaaact agttctgaaa 54360
ctatcaatag gcttgacact ctttagctgc caggagagct gaatctgaac acagggaacc 54420
ccacccagca ccccaaattt ggattattgt tttattttat ctttccccta cccccaagac 54480
agggtttctc tgtgtggtcc tggctgtcct ggaactcgga gatcctctgc ctctctgcct 54540
ctctgcctct ctctctctct gcctctctct gcctctctct ctctctctct ctgcctctct 54600
ctctctctct gcccctcgct gcctctctct gcctctctct gcccctctct gcccctctct 54660
gcccctctct cttcccctct ctgcctctct ctctgcccct ctctgcctct ctgcctctgc 54720
ctcctgagtg ctgggattta aaggcatcag ccatcacttc cagcttcctt tatcatttta 54780
aaaagaattt cctatgtgac tactgtattt aaatcaccac acggccaata ctcccccca 54840
actcctccca aatcccctct acccactcaa attcttatct tgtattcttt atcattatta 54900
tacatatgtg tatatatgtg tgtgtgtata tatatatata ctatatactg ctaatgagta 54960
acatttagtg ttattcattg ttgcatgttt tcaatgtgct ttccaggagg ctgggggat 55020
ggctcagtgg gcaaaattct agctgcacaa gcctaaggac cagggttcag atccccaata 55080
taaaggctgg ctggacatgg tggcttgcct atgatactag catgcttgct ggaagcaaag 55140
acagggaatc cctggagact tagaatctca gaagtgatct gggctggaca gactagctga 55200
actggccagc tctgggttca tcaagaaacc ctacctccat aacataaagt gtgatggaga 55260
aaggcaccta atgtcaacct caaaccccta cctgcatgtg cacacacata catccacacc 55320
acacacacac acacacacac acacacacac cacacacaca cacacacaaa taaataagta 55380
aataaataaa atatttagct ctccagacca aatcttggtg aaacccatgc atttgcattt 55440
gtgtgtgtcc tacaaacact gaaggttaag aagcatgctc cttagtaatt ttatagcagt 55500
ttgcgtttcc agattgaaaa cagattctat aggctacaca gtgctaaatg gattatgctc 55560
agatacagat tgaaaaggat acagattgaa aagggtcggg gtctgggcca ggatgacggg 55620
ccaactgatc tttgccgggg cttgtccttc agggaagggt tacaggattc accactgggg 55680
tgtggcctat ctgctgttag gacctgaatt gcctggagtg tttctagttc ccactagttg 55740
ttgaacttta ccttgaacct ctgctcccag ggaagtcatc aggactctgc catccctgga 55800
gtctctgcag aggttgtttg accaacagct ctccccaggc cttcgcccac gacctcaggt 55860
aaggggtttg gatttggaaa gatgcaattg ctataggagg gactctgaag gcagacagac 55920
gcaccgcctc ctcacgttgg ctagtctaat ataaacatcg cggtggatgg tgaggataga 55980
ctccatgccc ttttgtgaag gcatttcctg gcatcagctc ctgacttcag acagtttcac 56040
ccatcagaca aattgcctgg tgttggagga ggaggtgagc agggccattc ccatcatttc 56100
tcctcagaaa tggaaaggca aggaaaacat gaggttcttc agacacttaa tccctgggac 56160
tgcaaaatgg tggtgccccc cctccacagc tgctcacggc ggggcaggag atgagggcca 56220
aatgaagcat agatctagct attttttttt tagtgccttc agtaaattta aaatcaaata 56280
agggaaggga cctagatctt tatgttatgg cattgttaaa agtgagaact tgtagccagg 56340
gtgtggtggc gcacaccttt aatcccagca cttgggaggc agaggcaggc ggatttctga 56400
gtttgaggcc agcctggtct acagagtgag ttccaggaca gccagggcta cacaggtttc 56460

tctatacaga gaaacctgtc tcgaaaaacc aagaaaaaaa agaaagaaaa agaaagaaag 56520
aaagaaagaa agaaagaaag gaaggaagga aggaaggaag gaagaaagag gacaacatgg 56580
tctaggggtc agagagcaga atctccaaaa acaccaacaa tgcctgctgt aaatgtatgt 56640
cgttgatttg gggatgttgg cctccagctc accatttcct gccttagcct ccaaagtgct 56700
aggattatag gcttgagcca acacatctgg cttacgccta ttgtgtgtgg aaggggagtg 56760
ctgagtgtgc tcctgtgttt ggtacttata tatgaatata tgtatatacg catgtacgca 56820
tacttgcatg tgaaggccag aggccaatgt cagctgctct catcttatcc tttttattac 56880
attgtattta tttgtttgtt tgtttgtttg tttcatcgta cgcatgcagc cactcatgag 56940
catgtaacag cacaggtatg aaggtagact tgcaggagtc agttctctcc ttctgtcact 57000
tgagttccag gaccacactc cagcccccag gcctgggctg taagagagcc atcttactgg 57060
tcctctactt tgtcttctga gatagcatct agactcacgg aacctggagc tcatctagat 57120
ttacattggc tggccagctg atgcatttta aggtcaaatc ttcattccat ccctacccca 57180
cttccactcc cagtgctgga gttcgggaca cctgccacca agcccagttt ttcctggatg 57240
cagaagctcc aaactcaggt tcccatgttc gcatggcagg cacattttca gttaagcctt 57300
ccccccagct cctttaccct ggtctctgaa tgggggggag gctataaatc aggctgctct 57360
cagacattag gtaggaaata gaccatatac atgaggaaag atattcacct gccccatggg 57420
taccaggaag tgatgtccaa ctcctctttt gcttatcagg agaaatgctg actactacct 57480
ctggtaattt tgatgttggg aggaacaggg acattcatag gaccccattc ctcgctggtg 57540
agagtggaga caggttttct gaagggcagg agatctgtgt agaaaagatg gatgctgttt 57600
tctgaaggga aatggaggta gagtcgacct gggagagagg ggaggtgggg gatggttggg 57660
aggaatgaaa ggaagagaga cggcagttgg gatgtattgt ataagagaag aatcagaaag 57720
aaaaaaagaa aagctacctg caccccttcaa gtgttcctct gtgtgggagg ctgtctcagg 57780
gactacatgg gcaccgagag gcatcagtga gggtaggtac ttgatgttgt gtccctgaaa 57840
acaaggacag gaaatctgct gcatggccta agatggcaaa atgtggcaca atcaagtaag 57900
gcccaggatt ctgtctgtgg tgcagacctg ctgtagaatg agctcccagc attcccactt 57960
gctgtgtgga gacagcatgt tgcagagcca tgtgaggatg agggtccagg ccaggaggat 58020
gtcaacccac caccatgtag ccagtgggct gggggagctt gggcccacca aggagcttga 58080
gcagactgac agtgggttat gtacacaagt gggcgtgtca cacaaccgtg caacacagag 58140
aaaatccctg tgatgacaac ttctaaacca ccctgaggca aaaggagtag acaggggatt 58200
agagcctagc atattggagt cgagtggcca tgcagctctt ggaagcgtga ggaaggaaat 58260
ttcctggaag gataggttgt cttcctagca gcctcgtcaa tagatgtcaa tgtatgaggt 58320
agtacctgct acaatcctgc ttcttcagaa gactgaggca gggggattac ttgaacccag 58380
aagttctagg ccagtatgga caacatagca agagtctgat ttaaaaaaaa aaaaaagtaa 58440
agagggaaac caaataggtg acgtgccaca ctagtgtctt cctgctccaa gggtcctggg 58500
cacatgagct tgcttagtgc cagaaaagtc agaggaggag agggcagaca gagaccctcg 58560
tctccacctc ctttgactga ctaatggggc tggatataat ctgttttaca aaaggacagc 58620
ttttcagagc tgtttctatc taaggttgct ctgaatagcc atctcgaaat atgccagaga 58680
agaatattta gaggcggcct attttggtct cccacaaaga tttcacaggg aaaatgtatt 58740
tgtgttctat ttatacaact aaaaatatgc atcagcccgg ggaaactggc tctttgctgc 58800
ctttgaagtg aaggatgggt taatttctaa gaaagtaaaa gcaaatgtag tgcaggcacc 58860
acggatgctg ccagacacca gcgtttaagt ggcttgaatg gaagagcaca ccccaagtat 58920
ttgaagtagg tgaggagaga gaccaggtct ccagagttgg gcctgcagtg gccagggtaa 58980
gtccaagggg cagtatgacg cagaacagag tggcaacctc taccagtagt agaattcagg 59040
tctcattcct aacctcccat aaagcagaga atattgcact ctctttctct ctctctctct 59100
ctctctctct ctctctctct ctctctctct ctctctaact cacagagatc ctcctgcctc 59160
tgcttcctca ttgctggaaa taaaagcatg tgccaccaca tgcagctctc cttgtttgga 59220
gagagaaagg cagacagaca gacagacaga cagacagaca gagtataatg tgtgcagatg 59280
tccacagaac tccgaagagg gtgttggacc cccagaaacc ggagttctag gcagttgtaa 59340
gctgtccagt gggtgttgcg aactgaactc aagtcctctg gaaaactgga aagtactctt 59400
aaccactaag ctatctgtta gtcccccaag aatgtcttat cttgataggc cttcagatct 59460
cacagtccag ggggctgacc tgctacaagg ggtagaggaa agaaaactgg ctccaggcct 59520
gagttcagag tgacatctcc agagttcttc cttcccttgc cgtgtagtaa tttcctaccc 59580
tacacgagga gaaaaggaac agatatgtcc agtatgcctg gcatcttgaa agggcactga 59640
ctcttgccgc tgtagagccc ctttccttgg ggagtgcaga gaagtgctgc tagagaggtt 59700
caaaagaaac acaacagtca aaatagttgc tggccaaggg agggcatgtg ccctgtatcg 59760
ggcctgcaaa gccctttcaa cagtgtagcc aaggccatgt ttgacagtac aagcctgtaa 59820
gcccatgcat aagcaagggc tgggataaag ggctactgtt caagttagtt atatacacat 59880

caagtttgtt catcttacat ccttaggtaa gggtgggtag cttgttagct ccacctctcc 59940
agcaggaaag catgtccacg gtaaggtaga tactgtagag tttagctttg ttttgacctt 60000
tctttcttcg ttgcttggga ttgtacttgg gaccttcact catgcagccc atgtttcagc 60060
cttgcataaa tctataaata tatggttgga gtcaggtcca gcatatgtgg gtaccagtcc 60120
cagtctggta tagcaattta ttagctattt gtctttagac aagtaactat agggttctga 60180
gctgtaatgg ccaccaaaca ggtcctttaa gtatctttat tggccaaaca ccagggttaa 60240
aacactgaac acagtatggt ggtttcaact gtcatggagt ccacattcta gcaggggaat 60300
tttacacaca actcacaaat gatacagacc acagcagggt gatttgacag agacaacgaa 60360
gggcttgcct gagctgcagt agtccgggag cagctcttga tgggatctga gatctagatg 60420
gcagagcgcc cttcctagat gaaatccaag cctctggtta ctcctcagca agagtagctg 60480
gaggaagacc agaagtgagg aggaaaaggc tcagagtgga gcatctaagc ctgcagcacc 60540
tccagagaca ggctttggag cttgcctgtg gttcagccta ctgtgggaag agaccagtga 60600
gggttttaag ctgtggggat gggtatattt tggaaggcga acaagaacac cagcagctca 60660
ggcaatgtag gagggcggga aggtgatgga tgatcgagga cggccccaga gattttaagg 60720
actgtgtagg tggggtgaga agcactgtgg gtgaggtgga gttagaggga agaagtgaag 60780
atctattttt ttggccatga ggaatttggg gtgcacacac acacatacat acacacccca 60840
tcatgtcctt gtctacagac ggggtgataa tggtactgga aggtagaagg tggaaagagg 60900
accagcaagg aaggtatcca gtgcccaccc cacagctcac cctcagacag accttgttct 60960
cattttcatt acccaggtgc ccggagaggc cagtcccatc accatggttg ccaaactcag 61020
ccaattgaca agtctgctgt cttccattga agataaggta ctcacaggtg tcctgtagca 61080
cctcctgatt tctctgcagt cccctatttg cactgagtgt ccaagtccca ggttctgcat 61140
tcctatgact cgctgacatt cctgctgtcc acaggtcaag tccttgctgc acgagggctc 61200
agaatctacc aacaggcgtt cccttatccc tccggtcacc tttgaggtga gtgattctgg 61260
agcagtattt ggccgtcggt tcctgtgaag agagtgcttt tgatttaggc tgtgaggagc 61320
gaggaactca atggcagaga atagtagtca ggccagccac ctcgctgggg cagtgggtaa 61380
acactaaaat cgctgcctaa atgcaccagt gaagtacccc tcatgggagg ctgcaacgaa 61440
agagctcccc attggtccaa gatggtatga acctccttaa aataggaagg agaagtgaca 61500
tcaaactgag ggccacacgt tgcctataaa atggtgtctg gtacaggaac agtgggagaa 61560
aacaattatt tttgttttgg gggtttggtt tttttgtttt gttttgtttt gttttttttgt 61620
ttttttttt tgagacaggg tttctctgtg tagccctggc tgtcctggaa ctcactttgt 61680
agaccaggct ggcctcagac tcagaaatcc acctgtctct gtctcccaag tgctgggatt 61740
aaaggcgtgc gccaccacca cccagctttt tagttttttg tttcttgttt tttgggggag 61800
aaaacaattc ttatgatttc cttaaaaatg actggcttaa gtggagacct agaaaggtgg 61860
ttctttctct ttctttttct ttctttcttc ctttctttct ttcttccttc cttccttcct 61920
ttttttttt ttttttttgt ttttcgagac agggtttctc tgtatagctc tggctgtcct 61980
ggaactcact ttgtagacca ggctggcctc gaactcagaa atccgcctgc ctctgcctcc 62040
caagtgctgg gattaaaggc gtgcgccacc acgcccagct ccttctttct ttttttaaag 62100
gtttatttat tattatacat aagtacactg tagttgactt cagatgaacc agaagagggc 62160
gccagatctc attacagatg gttgtgagcc accatgtggt tgctgggatt tgaactcagg 62220
acctctggaa gagcagtcag tggctcttac ccactgagcc atctcgccag cccaggtgta 62280
tttttcttaa tggaccccca tgattcttgt acttccagct caacttggtc ctgcctttat 62340
gagacacccc actgcccacc ccaaagccct cttcatccct gagaggagtc tcacaaaaaa 62400
cagatttcca tttcttctta ccaccttggt tttttaaatt tcacatttat ctatcatctg 62460
tctgtctgtc tgtctgtctg tctgtgtgtc tgtctgtcta tcgtgtgtgt gctgagtgag 62520
tgagtgagtg agagagagag agagaatatg aaggaaaagg acagcttgca agagtggatc 62580
tctcctaccg tgtgggtccc aagagattga ctccagattg ccatgcttgg tggcagattc 62640
cttttcccac aaaaccatct tgcctgactt ctctactgct cctggcttta ggggggcttt 62700
tttgtttgtt tttatttttg ttttcatgag ataggacccc caaatagcct ccttgagtct 62760
acctaaacct tccaccacac ccagcttagc ttttcatttt ctaaacatgt ctgcgtagga 62820
atcgtcttcc taattattac cagtcgatgt tcaattagcg gttgtaacaa ggccccagag 62880
tgtgaatcta ttgctaagat tctttccgga agtagatgac acatcagatt aataagtgtt 62940
gtcaaactca tcagtgaatg aaccatttcc caacagttca tctgtttcct ccaccaaatc 63000
tgttggtgga gcagctgggt tacacatggg cgagctgggc tcagccgcgg tgccagccat 63060
ttttccagca tgctatggaa tataggggtg ttatcaggcc agggtgggaa aagcctggga 63120
atagcgagcc agatagcctc acccgggaca ttgaccatcc cagccatgta gataaatcag 63180
agttgacatc accaaagcca gccactgact gcatgtctta tgctctgaag gtgaagtcag 63240
agtccctggg cattcctcag aaaatgcatc tcaaagtgga cgttgagtct gggaaactga 63300

tcgttaagaa gtccaaggat ggttctgagg acaagttcta cagccacaaa aaaagtaagc 63360
cacccacccc aaccctgcaa cacacacccc atgttctggt tcctgtgtga gagctcttat 63420
gaatgaccag tgactagatg tcaacagcga ctgtcttagt tacttttcta ttgctatgac 63480
aaaacaccat gaccaagacc acttacaaaa taaagtgttt gatgtgggac tcacagctcc 63540
agggggttag agcccatgac catcacaatg gggagcatgg taccaggcag gggagtgtgg 63600
cgttggatca gtagctgaga aactacgtct tggttcacaa gcataaagca gagagggcaa 63660
actagaaatg ccgtcagctt ctgaaatcac aaagccccgc ccccagtgcc acatctcccc 63720
caacaaggcc acacctccca atccttccta aacagttcca gcagttggaa accaaacatt 63780
caaatatatg tgtctatggg agccattctc attcaaacca cctcagtgat tgtcagcaaa 63840
agccagtccc acccccttg cagggtagtc cctgtccatg ggtgctcccc aagccgctcc 63900
ctgcctatgg aagtcttttc ctgatacaag ggcactcctt aggggcctcc cctgcatact 63960
tagccgccag ccctgaaaat gagattataa atttgtcaca actgccccat agcagtcctg 64020
tgtacacacg gtgatttcat gtgtctgtgt gactgtcctc ttccctgcaa gcatgcagtt 64080
acctcactgt gccgggacag gccctgtact cacggtattg ttggcagagc tgctgggtta 64140
ccacccatag caaatcacta accctgtgga gctgaagtat atgaagcctg cgtctaaatg 64200
gtccaaactc tcactgccgc tctgtcactt ggcagctggg ggaccttgac ccgtctctct 64260
gaacctcagt ttccagttct cagggatgag aaagattgag tgaacaacat gagggacagc 64320
ccaacaccca gcttcctttc caccagctct cctcatggac accgttgagc cactgcagct 64380
cttctgtctg ttgtactgtc cagtatttgt acacacagca ggaaagttcc ccatgacagc 64440
tggcggagaa gctgctgtgg gcagagcagc atccagagaa gcaagccacc catatggtcc 64500
ctgcccacat gcaatttgta ttctggtgca ttgagggtga ataattctca gctaaatcta 64560
ccaatacagt ggcttcagct actcggaagc tttgaagaaa tcaggacagc atgataggta 64620
aagatccttt gggaagagag aatggacaga ggaggctagc ccttttaagc aatgagcttg 64680
tgagctgaga tctcaaaaaa gctactggtt ggacaggtgc ctgagggaag agtctcccag 64740
cgagtgcata aggtggagtg acaggatctt aaagaacaga aatagatgtg gggacagtgt 64800
agaatggaat aggagggcaa tgggaatgtt gaatgaccta acctcctgaa gaggactttg 64860
gaaaaactcc caagcaagaa gttttaggct agaccttggc gctgttgaat tatggatgca 64920
ggtggttctg agggtagccc tgggtgaatg ctcatctcca tatggggttg gacaagccca 64980
ccttcaaatt caggttctac ctcctgccag ctgtgtaacc tagcctctct gcgcctccag 65040
tctccatctg taaaatgggg atgaagacaa cagatgtcag gtggagattg catgcttatg 65100
agagcaactt gcacatctcc tcactcctct aagttgagtg agagaatgcc caggcagggg 65160
ctagaaaggg gacctgttga tcctacacct cagcactggg gaagaaaagg gctctttttt 65220
tccctcttgc attttgagtt cgtggcagaa gccaggacag attaccaaaa gaaaggagaa 65280
tggattgatc acatgtaagt ctcgtgacgt ggaagccctt acagggaaag caggaccagg 65340
aggatggcta agcagagtgc tgtagccaga tttccaaaag gtgggaagct agaaagtgat 65400
agcagagagg ggctgggggc taagtttagg gcaatgggaa gagcttggcc aggcccgtaa 65460
gctcagactc ctctcagtct tgtttgtcct tggcaataaa tgccttccct tcgggtacag 65520
agagcatact tgtcacatga gggttttatc tccagcttag aggaagacca gaaaatcctt 65580
tctccctcct catccttctt gtcagggtgc tatattttgg ggtactgggt cctcagctcc 65640
atcagtaaat agcacaggag gtattagaag tcaagaactg agtttcccag ctcagtcact 65700
gtgacttggt ttggtgccca ttctggtcct tggttgcatg ttgctacaag gcttaaaaga 65760
atgaatatca aaatgtgtga tgcctggcgt ggagagagtg tttggtgcaa ttggccactg 65820
gcggaggagt gtcattggta caattatcct tccatgtttc tgtgaggaca gaatggaatg 65880
ggggtggaag gtaaagagag gaagtaggaa agacagcctc tggactctgg agtagcaagc 65940
tgtagggaaa ttacagagct gagcttacat catcaggaca gggctctctc tgccttactg 66000
tgtctagttg gggagtgatt gatggtttga gcgggaaggc tatgccagaa gccgttggct 66060
tagaatagtg catactgctt gtacagggca aataagtcca gagattcatt gcagcagggc 66120
cactggtgac agctacttag agcaaactga actcccgaga agctctgaga gctagggttt 66180
ggtgctagct gggtgctctg gctactggga tggctgtgtg atgcactgcc cgtgctacat 66240
agatttaccc attccactgc gtgcccatgc ttcagagcct cctgtgtgcc atggtgttta 66300
tgtacagctt tctctgtcaa attaaaataa attaataata aaggaacaat gttgagaggg 66360
gagagagaga gagagaggat gtggatagat attttggggt ttttttttctg ttggttgttt 66420
tgtttcaaga cagaggctca tgtttcctat gctagggttg aattcgctct gcagctaagg 66480
tgagtcttga cgcccattga tcctgggtaa ctgcattcct tacataatct agatctcatt 66540
ggttttagga aagcaggaaa tacaaatgag cacatgttaa ataaaatcat atgttgtact 66600
gtactctccc agacagcctt gtactgtact ctcccagatg cttttaaaag taacaatgct 66660
aggccacatg tggtggtaca gccttttaat cctagcacac agaaggcaga acaggctggt 66720

ctctgtgagt tcaaggccag tgtgatctac atagtgagcc ccaggccagc cagtgctgca 66780
taataaagac cctgtcttaa agaaacgggg aggaagggga ggtgggagag gagggaggga 66840
gggaagaagg aaaaagaaaa atagtacaaa ctgagtgtta tggtacatgg ctggaatcat 66900
aacaccaagg aattttgcac ttttcttttt ataaatttac attgtgtgtg tgtgtgtgta 66960
tgtgtgcaca tgtgtgtatg cacttgcctg tgtgtgctca cccatgtaaa catacatgca 67020
aatgaccaga agaggctgtg gatcccctag agctggaatt acagttgtaa gctgcttgac 67080
atgggagctg gcaaccaaac tggtgtcctc tacaagagat acatactctc ctaaccactg 67140
agcctcccag cccacactcc aagtgagggt acacagtgtt tcttcttgtg tcacaggaaa 67200
ggaccaatgc tgagtgtttc agacccacaa aagccagaaa tgtcatgatt cagcctcata 67260
gttgcaaatc taagctgtcc tctcttcagc tatgtatgtg tgtgtgtatg tatgtatgta 67320
tatgtggctt ttttatgaaa cctatcttta attggaaaag tctcatcttt gtgtctttat 67380
tcttatgagc ttgcacccat gaaggaagag tgggaccagg caaaaatgag tgagaagaca 67440
aaacacacga tgggatgagt tacacagata aagtgaaata aggacatgag cctatgaaat 67500
aagctggaga gacggcccgg cagctaaggg cactgggtgt tctttcagaa gacccaggtt 67560
tggttccccc cctcaagaca tgacagcttc cagccatctg taaccctagt tctattggat 67620
cacatgccct cctctgacct ccatgggctc tgcgtgcatt cagtgcacag atatgtgcag 67680
gcaaacagct gtacacataa aataaagatt tgaatcaaag ccagaaatcc ctttaatccc 67740
agctcttggg agacagaggc agggggatct ctgtgagttt gaggacaggc tagtctatat 67800
agagaattca aagacaggaa tatgtagaga gatgtatgta tgtagtctca aacaaaacaa 67860
aatagaattt atttttaaga tttattatat gtaagtacac tgtaactgtc ttcagacaca 67920
acagaagagg gcatcagatc tcattacaga tggttgtgag ccaccatgtg gttgctggga 67980
tttgaactca ggacctcggg taagagcagt cagtgctctt aaccactgag ccacctctcc 68040
agccccaaaa tagagttttt aatttaaaat atttttgaat taataggaag agggaagaaa 68100
aaaaaaaaag agttctgtac aagggttaaa ggcagggcat caggaattca agccagccgt 68160
ggctacatac aaccatgttt caaaaatcca aaaagaaata aacagatttt caaagtaaaa 68220
attttaaagg actggagata catctctgtt gtccagtgct tatcaatcat atgcgggccc 68280
cttgtactgc cgaaataaaa taaaaacaac cagaaattaa aaacttgagg gaaggaagct 68340
cttagaaggt gggtttggtt tgggtttttg agccagagtc taatgctacc tcggttgcct 68400
ggcagctttc ccacaagagt catatggagg tactcaatgc ccccacctct gtcagtcatg 68460
acaagtctac cagggacttt gaggagggtg ctcaaagtgg ttatttgttt gtttagctaa 68520
ttagttttgt aatcacagta atcaaaccca aggacctatg catgtcaggc aagtggtagc 68580
caaggaactc actcctcatc ctcctgcctc tacctcccaa gtgctggtta tggcgtctgc 68640
cagtaggcct agctccaggg aagggcttct caaaacaaaa caaaaacact ccaatagcaa 68700
aaaatgacaa atgggatttt ataaaattaa aaaactctga ccagcaaggg aaacagttag 68760
cagagtgaag agacagaatg ggagaaagtc tttgccagtc acacactgac aggggatcaa 68820
tgccagtgat ggcggcacac ccgcgtttat cagagcgaca tttgttatag ccaagttatg 68880
gaactgacct tggtgcccaa cattgaagga agaaaatgtt cctgaacatc atgggaggag 68940
tattactcag ccataaagaa caacacgata ctatttacca aaaatggatg ggactggggg 69000
gtgagggggt acaattaggg aagaaaggga ccaggacaga aacaaagagt aactgtggaa 69060
tgatcaaaga gtgaatgaag aggttacaat gaatagatat gttaatatat gctcatttaa 69120
aagtcaaaca tgttttagaa aataattctt ggaaagctac tagagatgtt accagaatga 69180
aggtgtaaac aaagagagag ggagacagaa tgagagagag aagggtccaa ggagagggac 69240
aaaggagtag atgccagggt gctgaaggag gaaggtttcc aacaggcctc aaagtcagca 69300
gacctagagg ctgtgggaaa gcacttactc ataggacaga cagcaaagtg cttagtcatg 69360
tggggaatga gaggggagtc tgaatggagg aaagcacata gcagggggtt aggagggtag 69420
cttagcatgc agtgtgaccc tccgtaccac caaaaaagaa aagaacatgg tcattgtaca 69480
gttataaacc caggaaaggc caatataata tgccttttct catactgccc atactcaccc 69540
catactcctc cctctaactc ctctcagacc caccctcacc ccctccctcc tcccaacttc 69600
ctgtcatctc actgagtccc tgtgatggtt tgtatatgct tgggccaggg agtggcacta 69660
ttaggaggtg tggctttgtt ggagtctgta ccctgtacca tcacagtggg tgtgggcttt 69720
aataccctac tcttagctgc ctggaagtga gtattctgct agcagccttc agatgaagat 69780
gtaaaactca ctccttctgc accatgcctg cctggctgct gccatgttcc tgccttgggt 69840
gataatagac gtaacctctg aacctgaaag ctagccccag ttaaatgttg tccttataag 69900
agttgccttg gtcatggtgt ctgctcatag cagtaaaacc ctaactaaga cagaaattgg 69960
tatgtgctcc tgggtatgag accatcccct gggacctggc actacctgga gctacatcct 70020
taagaaactt gactttctct ccccagcagt catctgccaa tagttcctca gctaggggag 70080
gggctcatgg gcccaactca gggaaacgtt tttaagttac ataggaaaaa gaacaaccag 70140

gactggtctg tcaggagcct tgtcgtggtc acagtgacat aagcattgag tgtttatgta 70200
gctaaaatac aacaataaac caggtggtca tgcatgcctt ttatttatgt ttctattttt 70260
taagaaacat gagataaagt taaaagagtt cattgaaaaa cttgaaatta gggctggaga 70320
gatggcacaa cagttaagag cactgactgc tcttccagag gtcctgagtt caattcccaa 70380
caaccacatg gtggctcaca accatccata atgaggtctg acaccctctt ctggtgtgtc 70440
tgacagctac agtgtattca tatacataaa ataaatatat cttttaaaaa aaaaaccttg 70500
aaattagttg ccttcagggg gcaggacatt ggagggagca gagtggaaag atactattgc 70560
caacaacaac aaatcttaaa gaataatttg attcattttt aattttatcc ttattatttt 70620
taattatgta tgagtgtatg cctgggtatg tgaggtgctc caagagacca aaggagtcgc 70680
ctggagctgg agttccaggt ggttgtgagc tgcctgatgt aggtgctagg aattgaactt 70740
ggatcctctg aaagagcagt acatgttcct agctgctcag ccgcctctcc agcccctcaa 70800
tttttgatct aaaaaaaaag ttgaaagcta caaattttct gtgaattttt cacaaattaa 70860
gcaagacaga gaaaaggtag ggaggtgaag tggcatggaa ggagaaggag aaaaggatgg 70920
acagtctcat ggtgtgtgga gcaaagctaa ctctagctgg ttctgggtac cacactttaa 70980
aatgagcccc cagggcttag cccaggtatg tgggggccag attggggctt gggtgcctcg 71040
ggagcaatgg acagtggaca tagcggataa agaaacaaat gcatgaatcc ataaacttct 71100
gtttttctga agtcctgcag ctcattaagt cccagaagtt tctaaacaag ttggtgattt 71160
tggtggagac ggagaaggag aaaatcctga ggaaggaata tgtttttgct gactctaagg 71220
taagtgacaa tagacatcta atgtgaggtg aaaaggatca tgggccttga ggaccgtgct 71280
gtgctaccag ggctggcgat cattccctcc ctgggttcta cagcaacctt ttcattcctg 71340
ggaagctgaa catagctagg taagaaccag gctctgggcc agctgggact tgcagagttt 71400
tagttccagc gtcccaggat cacaagtact atcttccagg aacaaggaga cactccttgt 71460
ctcaggccca gtgctgagct ggccaggatc ctagccttta ttaacagtgt ctcttggcga 71520
gatttaggag ccctttgggt aagctctgga tttaagacct gaaccagtgc tcccctgtca 71580
cttctgtgtg actttctgttt ttaaactctg ccctgctctg ctaaagtaac ccctggatgt 71640
gcttcatcca atccagatgt gtggtctctg catcttgatt gctcacctca gaagtcccct 71700
caaatgggct gattaccaca cagaggcctt gctgttcacg ctcagtgagg cgctaaaagt 71760
ggatagaaga acagatagtg tgggtccaca tgctgcccca cctctggggt gtgggacatg 71820
ccacttagga aggtgaaaca caggagtttg cctgtgctta tcccatccta ctgccatcag 71880
gtgctgggct tcagggaaga accaagacac cattctgggg gtaggaagat gcaatctaag 71940
ataggggaga ttgaagttgg ggtttcctga cttccgggcc tccagttgct tctggagaac 72000
tgcacagaga agtcggtaat gaagtgtcag ggattcatgg gcgacacagc gtgataaggg 72060
ggaggggggg ctgtgtgtgt gtggctgagt tgcaaggaat tgagagcaga gttcagggag 72120
ggacccacag gctgatgact gggactagcc caggggggca ggggaatggg aactctgggt 72180
tagcttatca acaaaggggt gcagatggac tttctatggg agacttaagc ttcggctcta 72240
tagatgaagg ccttccccat gctccccatg gcagttctca atgaaagaga cagtccttgc 72300
ttccaaagtg ttattggttg ttcatcatgg aaaatggctg cttaccaact tctctgggtg 72360
ggccagagat taaatacctt cttcaagaag gaatgtcggt tcaagaaggg acactttgtg 72420
gctaaactct cagggcctct aggtacctac ctcattagta tagagaattc tgttttgggc 72480
tatgtgaaca acatggggaa tatggtatgt gttccaggcc aggaatctgg cagggagcag 72540
ggaggcacct accatcttaa gtgtgagttt ctggagtctt ctgttctcca ttaccaagtt 72600
tcctggcatg gtgaactgtc ccacaagata gggcacatag cttagcatgt gtggctgact 72660
cacgggtcag acattggaga caagctaagg gaaggctccc agtacagtct catagaggta 72720
acccccacca ggttcccaag gcatctactt tcactcatcc tgggagggaa tatgaaggta 72780
cagtggcagg catgagccga cgggaacagt agcaagagag tacagaaagt agagaatgtt 72840
cacatggtta ggcagcattc taggcttcag gagcataggt cccaaactgg ctatttctgt 72900
cccattgctg caagctcagt ggggttgaga cactgaccct tgagccatcg ctggttaaag 72960
gcaggtttca ataaagaagc tcttgttatg tcatgatgag ctggtatctt tagagtgggc 73020
cactgtgata tttgcaatga tgaggttgaa tcatatgaaa ttgctggctt tgaaaccagg 73080
tgagggacac caggcatccg gcctttcata ttggcacttg tcatagttag ggttttactt 73140
ctgtgaacag acacaatgac caaggcaact cttgtaagga caacatttaa ttggggctgg 73200
cttacaggtt cagaggttca gtccattatc atcaaggcag gaacatggta gtgtctaagc 73260
aggcagacat ggtgcaggag gagctgagag ttctacacct tcatctgaag gccactagga 73320
gaagactggt tttcaggctg ctagaatgaa ggtcttaagg ctcacactca cagtgacaca 73380
cttcctccaa taaggccaca cctatcccaa caaagccaca cctccaaata gtgctattcc 73440
ctgagtcaag cattcatagg actgcttgaa attttgtgtg ttctcattta aagtcagcca 73500
ccctattgac gacagccctc tagactaaga gagagacgag tctcatgctc cgctgtgagc 73560

ataggagaaa tacatgtaag agatcaggga ggccctttc ctaacatgcc ctttggcctt 73620
tgcttagttt tatttaaaaa aattatttta tttggaggtg gatgcatgtg cctcagtgca 73680
tgtgtggagg tcagaggaca acctgcagga gttacttccc tctctaccat atgagtccca 73740
aggattgaac tccattgcat agtgtatgtc ttcactgcgg agccattgtg ctggccccgt 73800
ttctctacat tttctctcac caaataagta attcagtctt acagaaataa aataaaaacg 73860
ttgcaaccaa aatagttaaa agtttaaaag aaaagcagca cggccttcac tgctaaccgg 73920
ttctggactt ccccttccgg tcggcttaag ggattattgt cgtcatcgtc gtcatcgtcg 73980
ttgtttgttt ttatggtttg ccttggtttg ggctttcgtg agacaggatt tttgcaatat 74040
tgttctgtct agcctggtgc tcactattta ccacagattg gccttgaact cacggcagtc 74100
ttcttacctc agccttccaa gtgctggaat tacatggcca tgccccacac ttttttttt 74160
taaagattta tttattaatt atatgtaagt acattgtagc tgtcttcaga cactccagaa 74220
gagggagtca gatctcgtta cggatggttg tgagccacca tgtggttgct gggatttgaa 74280
ctccggacct ttggaagagc agtcgggtgc tcttacccac tgagccatct caccagccct 74340
ttttttttt tttattgcaa tggtgataca cattccattt tgccaagtgg cattttataa 74400
aaattatgta ttgtggccac ctttctgtgg ataaagtaat tctacaccat aatgttaaaa 74460
ctctgtgtgt gtgtgtatat atatatacat acatatattt gtgagtccct tacctgtgaa 74520
catgctccag tgctctaatg tatcttggtg cagttcctca ttttttgctt caaatcataa 74580
ccttgcagtg gaacagctgg gtccaagaag ccatatactc atagctttga atgcatctct 74640
ccgggtgggt gcgtctctca tttatacaca cacacacaca cacacacaca cacacacaca 74700
cacacacaca cacacacaca aagacctttc agtatttttt taagctgacc acatctttgt 74760
agaggccagt gtgcagggag gtggaggttc tgctaagtgt gtctgtcctt cctccctcag 74820
aaaagagaag gcttctgtca actcctgcag cagatgaaga acaagcattc ggagcagcca 74880
gagcctgaca tgatcaccat cttcattggc acttggaaca tgggtgggtc cctgtgcccc 74940
ctccatccta ccagctctac ttgggtccac cttcctgcct cagcttctac aagtggcaca 75000
agggggcacc tctatctctc agccatagtc ctggtctcaa tcccatctat acaacctctt 75060
ttctagagaa ttctaccttt gtcagaggat gatgaacaag taaaataggc tttaggatta 75120
gtgcacccca aaagtagcca aagtagctta ctctgcagtt taccaagggg ccccaggcct 75180
ggaataaaga agtctggtcc ttgcttatct agggaggaag tgaggggagg gatagagaca 75240
tcagtggatg agtagatgga tggatgggta agtggattga tggatagatc ctagtggccc 75300
taagtttaag tgccatataa ataatatata aggtttttt taaagattta tttattatat 75360
gtaagtacac tgtagctgtc ttcatacaca ccagaagagg gaggcagatc ttgttatgga 75420
tggttgtgag ccaccatgtg gttgctggga tttgaactct ggaccttcgg aagagcagtc 75480
gggtgctctt accaatgagc catctcgcca gcccaatata taagttttta gcatatatgt 75540
gtagttaaaa gtaaatgaac aaaatatgta tgctaaatat aatggtagca taatcatatg 75600
gtaaatatgt aagtgaaatg atcatttggt ataagtactc aacaataaaa gctaaggtac 75660
aaaatatata ttggattaaa tatatatcat atataacaat taccacatat tattatacta 75720
aatatatata atagtacatt atactaattg tcaattatac taaatgtttt gggatatata 75780
ttgagctaga attaaatttt taaaatgtcc agttaaatct gaatctcaga caaccattgt 75840
atggggcata cttactctaa aaatcaccca gccaccagaa attcaaattt aatggatatc 75900
ttagtctatc attaactcct caggcactca ccctcaggtg ttcaggcatg tggcacaagg 75960
acctttctag ttctcttgtc cgtctgtgaa gagttctttt ggttgaacat ggcgggaggt 76020
catcccattg ctgcccttcc agcagcaaac acatgggggc gcccttccag gagaagtcac 76080
atggaggcac taggcaggag ccctgctgcg catggcaggt ctccttatga gtttcactta 76140
gagccccccg gggggggggg gggagtttgg cagagatcat tatgtggatt attggcttct 76200
tgtctctctg ttggctgagg gacatagata gatctcctca gggtgtgggg agaatatcta 76260
cccatactta gggtagctcc agaagtagat tacccatcat tccttgaggg agctgcttcc 76320
actcattttt caaaacggaa gctgagacaa gttgcagccc cctcctgttt ctctgttgag 76380
cacctcttca tctcagttct tagctgactt ataactttcc tccaggaagc ctttggtgtt 76440
tcctggctgc acggtagaag cctccatccc ggacatgtca ctcacccctg aaaatgtaaa 76500
ctcctgagtg acaaacgggg ccagatttca tgtctcactg ccaccctaga gcctaggaca 76560
tgtctggcaa cctgttcagt tagatcaggg ctccagagag gatctcccag gctttccatc 76620
ctttgaaatg taatgttatc tttacaacaa gatctcactg agctctgccc atagctaaca 76680
tgtaaggtgg tatcagcact gagtcctcca tctagggcat ttgcccccat gatttttgta 76740
actatgtgat tctcctcttc ccctgttttg aatcccccta atgcccttga cctgacctgt 76800
ctttgactca ttcggtgttg ggttttgctg ttgaaggtaa tgcacccccc cccaagaaga 76860
tcacgtcctg gtttctctcc aaggggcagg gaaagacacg ggacgactct gctgactaca 76920
tcccccatga catctatgtg attggcaccc aggaggatcc ccttggagag aaggagtggc 76980

tggagctact caggcactcc ctgcaagaag tcaccagcat gacatttaaa acagtgagca 77040
gctggccagg cctggggtgg gaagacagca gactctttca agcattccag aagtcagaca 77100
ggatacttcc aaagatgtat aggattgctc aggggtaccc cactttcaga gccacagatg 77160
tgcattgagg tggcaccctt acaagttgat agggtcctga gtccgccatc ttccctactc 77220
ctgcttaaaa gaataatatc gccgggcgtg gtggtgcacg cctttaatcc cagcactcgg 77280
gaggcagagg caggcggatt tctgagtttg aggccagcct ggtctacaaa gtgaattcca 77340
gggcagccag ggctatacag agaaaaaacc aaaaagaaaa gaaaaagaat aatatttact 77400
tcctagatgc attttcagtc ccagttctca tctctgaggt gctttgtctc atttctaggc 77460
attgttggaa gttccccctg aaagctagga aatacagaca gggtgtctta ctcccagggt 77520
ggaaccgggg tgcatgttca gggttccaaa ggtgtacctc agtcctgtgt tgcaacacca 77580
tctcccacca ccaccaccag gttgccatcc acaccctctg gaacattcgc atagtggtgc 77640
ttgccaagcc agagcatgag aatcggatca gccatatctg cactgacaac gtgaagacag 77700
gcatcgccaa caccctgggt gagcatagag ggaaagccat tcctgtgcat gctcctcctc 77760
ctcctccttt tggcaatgtc ccaggataaa ctgtgagagt cctgtcctgg gatcccttcg 77820
tctctagcaa atccagaagg ttttccttgc aaaaactcat ccagggtcta taccagctct 77880
ggcaatctgg ctaaaatgtg ggttctgtct cagtaagtct cagatggttt ccattctgac 77940
aagctttcag gtgacatagt gccagccagt ccagggatcc cacttggaga agtgtgtgta 78000
cttgtgtgtg tccgtatttt ggggtgtgta tatgtgggtg tttatgtgcg tctgtgtctg 78060
tgtatgtcnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 78120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnat gtgtgtgttt 78180
atgtgcgttt gtgtctgtgt atgtctctct gtgtgtgtct atgtgggtgt acatgtacat 78240
gtgtgtatct ggggtgaaca tttttgatcc cacaggagtg gatcctgagt agatagtcat 78300
tcactcagtt gggggaggag gcacctccca aatcccaagg ctccaggctg ctccacccct 78360
ccttgtccct ctctgctcag tatcatggcc cccaacatcc ccatagccag aaacagatga 78420
tagcattctc cttcctttct cctatgtaga aaatcacaag gctgctacca cctgccagtc 78480
actggcatgt cccacccccg ctgacctatc cactctcctc tcttcccttg gtctgcttct 78540
ctttctcctt gctaacagca ttagtaccct ttacctcacc tcagcctctg acccctcggc 78600
aagcctgcct gcctgcctga cattctgctt cttctctttc ttcctccctc aggaaacaag 78660
ggagcagtgg gagtgtcctt catgttcaat ggaacctcct tggggttcgt caacagccac 78720
ttgacttctg gaagtgaaaa aaagctcagg taatgggagc cattccctcc atgcacccag 78780
acagccttag cctcgcacac cctgatgctt gcccagcctc aggaccttct ggacatctgc 78840
cgtctgagca aagacaactg tagtacagta gagacctcag gtatatgtga ctcttgtctg 78900
gaggacagaa aacatgttga tgtcattatc aacctagtga acccctttga gcctggttca 78960
tttctaaagg gaaataccac tcaggagcat cctatagctc aaagctcact ttgtctgcat 79020
gtgcctctgt catgaccata accagctatt tcaattgctg atctgtttcc taatgggtag 79080
attcctgcat ggaatctaga ctctgtgttc ccagttgggt ttcccagaca cactcaggtg 79140
accactggga cagagcttga gtctaagtct ctgtttccca cctgttgcct ctgacttcct 79200
actgtcagat caaggacagt acccatgcta gatgtgctag atggcccttg ttttcttgaa 79260
ttaggagaaa tcaaaactat atgaacatcc tgcggttcct ggccctggga gacaagaagc 79320
taagcccatt taacatcacc caccgcttca cccacctctt ctggcttggg gatctcaact 79380
accgcgtgga gctgcccact tgggtaagga gactccacct tgggtcgagc cataggggac 79440
agaggctctc aggagtcggc tgtggacaga aaatcaaaga gacatgtgaa cactcagaac 79500
caggctaagg tagcagctct tcgcaggatc tcaatacagc tcctgtcatg gatcctctgc 79560
ctttcttcat agttcccttc tgcttgtgcc tcctttgccc caggacttct atggctgccc 79620
cttgtcactg tcactgtctt tccagtcagc cagccagcca gcagcagctg ggcttggtca 79680
ggaagcacag acatcttagg gggccagctc acttatctca acaaaccatg gggtcactgg 79740
tggatccaag gcattgcttg cctgcctaca ggaaagggac caacagatgt gaccacaggg 79800
atgagtctag aacccacaca gttgttttgc atgctcataa aagttgaagc aggagacagc 79860
gtggggtgga acagtgagtg cccttctgtc tagaaatcag agttagcagc ctgcttccct 79920
cccacgcagg gactttaatg tattcccaca ctcgagtcct gggagggaca aaggaagcaa 79980
agcaggagct gaccctgttt tatagtcaaa cacacagaga ggctagggaa gcggctgaga 80040
ggagattgca catcattcat agctgaggtt ggacctcagt tgtcccttat cccaagcatc 80100
tctccatatg acccatgtcc cttggtctcc tttgggcttc cagccaccat cacagctctg 80160
cagccagctc accatgcctt cctcactagg gacccccagg ggacatcttt cctccctggg 80220
aatgaacagg caggtaggag catgagaaga gtttccatcc taagcctcct gaggaaggcc 80280
tgccacagct ccacgctcac cagggcagaa gcccagagcc tggctcaaag ccacagagaa 80340
gtagacagaa gagagaactg ctggagccca aggccagtga attccccagt tctgcactgt 80400

ggcaccaatt cttgtaatta ggctgagtgg ttacaaagag ggtgatgctc agaagcctca 80460
gagccaggga ttccccacca ccactagtgg acttaacact agatgccttt gttcccacag 80520
aaagagccag agaaagtaat ggagaaatac tacctgtctc atgaagtatt gctaacaatg 80580
gtgtttaaaa gctgatgtat aaaaacaaca gatctacatg caaatttggg ctagtctgta 80640
gctaattatg gtgatggtgc agcccttcta ggctttcctt atccttataa agtgattgcg 80700
agtcaggaat cagcatccca gacggatgtg gtcattggga agtgtggtga agaacacgat 80760
ctaacctggc atgtccccgt tccaggaggc agaggccatc atccagaaga tcaagcaaca 80820
gcagtattca gaccttctgg cccacgacca actgctcctg gagaggaagg accagaaggt 80880
cttcctgcac tttggtgagg gcacactgtc tctcctttgc atcttttcct tccatcttct 80940
ctctacctgg acatcctgac caaggacagg gctgtgtctg cagaggaggg aagctggagg 81000
tgctagggaa gccaaactga tggccatgcg tctgtttcag aggaggaaga gatcaccttc 81060
gcccccacct atcgatttga aagactgacc cgggacaagt atgcatacac gaagcagaaa 81120
gcaacagggg tgagtcctcc cagaagccac tctcctgccc tgtcccacct cctttaccca 81180
actcactatt ccatggtggt tcctagaaag tgggaagtat cctacctcac cagacaacag 81240
caaaaacaaa agtcagaatg cacatacagg ggccagggat ggaactcagt catagagtgt 81300
ttgcttagtg tgcacaaagc cctgggttcc atctcagcat caagtccagc atggcaccat 81360
ctatctttga tcccaatact caggaactag aagtaggaag caagaagatc aggggttcaa 81420
ggtcatcctt gactacaatg aatttgaggc cagtctgagc tgcacgagac tttgtcctcc 81480
cacccccaa aagaaaactt tcacatgctg acttatatct tggtacaaaa ctgccaccaa 81540
cttgtacatt aaaaaataat ccaaaaagct gaattaacaa tgaccctatt ctaagatgct 81600
gagagaattt taaagcattg tttcttttgt tttgttttgt tttgttttgt tttgttttgt 81660
tttgttttgt tttgtttgaa acagggtctc actatgaaaa tcctggctgt cctggaactt 81720
actatgtaga ccaggctggc ctcaaactca cagagatcca tttgcctctg ccttctgagt 81780
gctggcatta aaggcatgca ctactatgcc tggttactag taaggataat tttatgtgta 81840
ttagtggttt tgcctacatg agcatctggt gcccacagaa gctgaaaggg ggtgtcagat 81900
ccccagaaac agaatcnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 81960
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnntgca 82020
agacaaccag agatttaagt actgagccct ctctaaaacc cttttaaaag acaaaaacaa 82080
agatacacac acacacacac acacacacac acacacacac acacacacac accccacaca 82140
cacagttctg ttagctgaaa ctaaacaggt tctcatgcat ctgagtgtgg tggtgcatac 82200
ttgtaatccc aatacctggg aggcttaagc aggaagatca taggttcaag gccagcctgg 82260
actacatggc tagactccaa aacaccacca gcaaaactga tttctttgct tcaggactta 82320
gaatttcaaa acattgtagt tctccattga taaagccaca ggccaaagta aaccggtcac 82380
agagtgactg ttaatatgtt aataaggcca ctcatgttga aaggaataat gtttggggaa 82440
agatgcaata aaatactctt gcagaggacc tgggttcagt tcccagcacc aacatggctg 82500
ctcacaacca actgtaaccc caactccagg ggtctaacac tctcttctga cctttgtggg 82560
cactgcacac atgtgttgca catgcatata tgcaaacaaa gcactcacac acataaaata 82620
aataaatata ttttaaatag tttagcacta tctgagggta aattccctaa agaaccaatg 82680
acagctgaac ttgatgatgg ccgatgtcct cattataggg tggttttgtt tttaataata 82740
acagagaaag cagcacatgt gctataaaac tgtcatgtta cattgcagat gaagtacaac 82800
ttgccgtcct ggtgcgaccg agtcctctgg aagtcttacc cgctggtgca tgtggtctgt 82860
cagtcctatg gtgagtggaa cacggtgggg tgcaggctag gttttgggtt ctgaggacag 82920
tagcaagcca aggggcttca gtctgcttct ccataagatg agcgagtgtc ctgaaagagt 82980
ttgtgcatct ctatcccccc tgggctgcag tgtaataaat cccgctcaga gagagacact 83040
ttaactagaa aggacttttt gttgttttct ttttcttttt caagatttat ttattttata 83100
tatatatata tatatatata tatatatata tatatatata tatgtatata catatgaggg 83160
catcggatcc catgacagat ggttgtgagc caccatgtgg ttgctgggaa ttgaactcag 83220
gaccttggac ctcaggaccc ctaatatcat aaacaggacc ttctggtact attcttttta 83280
agagaaccag actgagtttg agccatataa aagtgatatt ttatcaggta taaaacaagc 83340
aactaccagg tattcacatg gttcagccta atgcatatca ttaagtatgc ccatgcagat 83400
ttggagaggc ctaagaatat tttattgtgg ggctggagat atggctcagc agttaagagc 83460
acttgctgat ctcccagagg acctcaggtt ggttcccagg acactcactg tgtggtccac 83520
aacctcgaac ttcagctcca gaggatccaa ctgcctcttc tagatcagga gcatctacac 83580
acatgcacat gcacacgcac acacacacac atactttaaa agacaaaagg aaatcttaaa 83640
acacacacac acacacacac acacacacac acacacaaat gtccaggcca tggcactcaa 83700
ccctcatccc accacaaacc acaggtgagc aagtcataag ccagagacct ctgataggca 83760
gctcatgacc catgctgtcc agctgagccc tcatgcctac ccatccctga gagtagagaa 83820

actggttggg aggatttggc cactgactag catccacctt atggtcccct ggagcccatt 83880
taactccgat cccacagatg aaggtggttt ctgtcccctt catctgtatg tgtctgctgt 83940
tgccctggga tactttatca catgcattac aacccccctcc ccctgctctt gagctctgtc 84000
tcccagaagt ggcctcacta gataaacttc atatagcacc tgtagaggtg tgacacagaa 84060
gcatctctta cactcacagc cttggccaag gcactcttcc gccttcagaa cctgaaagtg 84120
ggacaggagc tatatattgc tcaggggtag agaattttcc tgccatgcat aaggtcctgg 84180
cttccattcc tagtacttcc caaaccaaac aaacaaaaac gccagaaaga tggtccatct 84240
cccatccttt accaggactc cttcctcagt gtttcctgct tgtctgcccc caccaacacc 84300
tgatctctcc agccctgtga ccctctttct agcactacac agagtcagtc atatggaaac 84360
ctactcccca aatcctgcag ccctgagggc cctctcgatt tctgactcaa aaggtcctgc 84420
caactatctc catcagtcag cgacctctgc cgagggccaa gtgataaagg ctcatagaat 84480
tgccactatc tctttagaca ccagccttgg ccttcagaac tgctccatag cacaggtttt 84540
aagaaagggg aatgcgatgg atcaggaagt ggcaactaag attgaggaga gtaaacatca 84600
cgaacattcc cagcatgtga gaggcaaaaa ggaaaaccaa ccttatatgt aaggctgtgc 84660
catcttccac aggaaggctg aagatgtata tattttgcta atattctttt gagagccaga 84720
ggacgcagta gaagggttta aggagttggc catagaggaa gccaagaagt tggggaaaga 84780
aaaaagaatt catatctaga atgcaagaaa gggggccctg ggacgggcgg tggtggcgca 84840
cgcctttaat cccagcactt gggaggcaga gacaggcaaa tttctgagtt tgaggccagc 84900
ctggtctaca aagtgagttc caggacagcc aagactatac agagaaaccc tgtctcggaa 84960
aaacaaaaca aaacaaaaca aaacaaaaca aaacaaaaca aaaacaaaaa caaaaacaaa 85020
aacaaaaaca agaaaggggg ccctgggttt tcttccaact ggttcataat cgggtgtatc 85080
aactatctat cgctgcatag caaatgccct aaatcccagt tcttaaaaca ctgagtcctg 85140
ggctggagaa atggctcaga ggttaagagc ctgagttcaa ttcccagcaa tcacatcaca 85200
tggtggctca caaccatctg taatgggatc tgatgccctc ttctggtgtg tctgaagaca 85260
actacagtgt actcatatgc ataaaaataa ataaacaaat ctaaaacaac aacaaaccac 85320
tgagtccttt gctcatgatc ctttggctgc actgaattca gacagctcat ttcttctctt 85380
tggtgtcagg tggacacact gaagtatttg atcagatctg gggttggctg actgtcagca 85440
agggtgatgg cgatgcccag gaggcaagtt caggcacctt tgcttcttgg actcagggtc 85500
cccagactag gtctcagctt gtccctacag tgagtttgtc atccacacct tggtcaaaag 85560
gtctcctcta tccctactta gaaatgcctt ctctcccttg caggcagtac cagtgacatc 85620
atgacgagtg accacagccc tgtctttgcc acgtttgaag caggagtcac atctcaattc 85680
gtctccaaga atggtaagca atgggcaacg tcagcttttc ttgttttcct caaagacaag 85740
gggcctaggg catttgtcat ctggttgcag ctaccaattg tctgggttag atgtaggcct 85800
atcctttcct tcccaaggcc catggtctgc cctacctgat ctcttcatgt tcaggccaca 85860
tcacaatctt agaccagaaa accatatata tatagtgttt taagtaagtt tcttctggat 85920
aagcaagtgc ttcaagtttc ttccaattgt ggtccaattt atttaattcc atctaagttt 85980
aaaaacggac tctaattaaa gaaaatatta gtccagtgtt agtgtgggtg catagtaggt 86040
tgagatggtg aaagctatgt cggtgggatg tggatgaagg aggggaggga aactggaata 86100
aactcatgct ttgagccaca ggtatggttg caaattgccc agcccagccc tggttgctga 86160
gctacccttt ctgcagcaat ggctgggtgt gcctgttaac aagtccatcc ggtcctatct 86220
taggtgacca ggaagccatg ctgtcatctc ctctacccac ttctgtgcag cagataactg 86280
atttagggat gctgggtatg ggtaacccaa atacagacag agaacagctg tctcactcta 86340
tggcctctgt gctgagggat gtagacgaag ctagacctct gtctttccaa ccattggatc 86400
aaacatctgt ggatctgatg tgacctccct tctccaccca agtcttcaca cctgtccagt 86460
ctccttccta catctggcca cctttatagc ctttaggctc cacccccttcc tgctcatacc 86520
tgtcccctcc tttgagtttt tcagcattga aagagtagcc tctatcacct ccctctttgt 86580
cgggcttggt ttccttctgc tgggattcag cacacagcag gtgcttgaaa atattggtca 86640
ccagtttgga gttttgatgg aacatttgtc taagcagagt ctggccaagg tgatgtgggg 86700
tttagaagga gagaaagaag actaagaggc atggtggagg ttgctgtcaa gtggcttaga 86760
atttatatga aaagatagaa cacttgctta aaaccagttc aagatcagtg tggtggtaca 86820
cttaggtggt ggttctagcc cttaggaggc agaggcagca ggctcagaaa ttcagtttga 86880
gtctgactaa actacataga gtttgaggcc agtttgggat acatataaca ccctgtccta 86940
caaaacaaaa acaatgataa aaaacagatg ctgtgtaaat aaaataaatg tcatgtaaat 87000
tagcatagtt ataaccatga tagctgaagg ccagtaagtc ccagattcca gagtagagac 87060
tggttataac tgtgtctgag gaagtaagag caggtctccc tagctgtgga ctcaaaacag 87120
cagatttggg tgttgagcct ctaactctct gccagaactt gtggggaaat ttggttgtat 87180
tagtgcatgc ctttaaccct agcagtggag cttaagaact tgatgcagga agatcccaag 87240

tttgaggcca gccttggcta catagtgaat tctagacaca cctgttctat atagcaaaaa 87300
ggagagaggg acagagggag aaggaggaag taggaaggga aagacagaca gacaacagag 87360
acaaagagac agacgggtgg ggggagggag gaagagaatt aattagttgg gagactagag 87420
tgaggttgag aggttgcccg aggccacaca gattattgtc aaagctggga ttgtagccat 87480
atcttgggtc cttgctgtct ttattcttcc tgacctgcac cctacatgaa cagcttgtga 87540
cagaccagag aatctgtggc catcaacaga gatgctgagc ttgggtcagg aagctttcct 87600
ttaagacaaa tctctccttc caacctggac atttcccatc tgtgggccct cagagggagc 87660
ctggcagagt gggtgcacag agaggaaaac cccagaaaga agctggaacc tccatcttca 87720
gttcaggtac agtatacaaa taaatcctag tctcccatgc tccaggtcct ggcactgtag 87780
atagccaagg gcagatcgag tttcttgcat gctacgccac actgaagacc aagtcccaga 87840
ctaagttcta cttggagttc cactcaagct gcttagagag taagtgcctt gtgaactacc 87900
ctttggggag gttctttcac ataactacct ttccattgat aactggttga tccctcaccc 87960
agttcattcg ttttctctct ctctctctct ctctctctct ctctctctct ctctctctct 88020
ctctctctct ctctcataag aatccatgga tggacccctc agtgggttca gcaaaggtgg 88080
ggtacctatg gggctaccca gtggctatgt ggccctttcc ccacccctgt ggtttgttct 88140
tctactgcct cagccctaga gatagaggca gccactggtg tccatgtccc tgtcccctag 88200
gtcccccctg tccctgactt gtccatcaac attctctagg cagaatccat ttgagtttcc 88260
accaagctgg gtttgagttt ccattaagct gggtttgagt ttccaccaag ctgggttctt 88320
cttgctcagt gaaagccctg ggcaaagctc cagaatgctg aagatctaac agggcctaaa 88380
tagctgctac tgcctgcttg ctgttgccct ggccctcccc ttgccctttc catattgtgg 88440
ctcttgtgtt gtggggctca ctgtgactgg gagcaaggta aagatagaca gcccatttca 88500
ggtggcaacg ttgagcactc accatgctta tgaaagatcc tttctataga acatctccag 88560
acatactgtc tgagactcct atgcaagaaa actatatgct ttctaagtct tcctgagatg 88620
gttcaaagag tgatggggaa ggataggaac caaccttgca aacagaggca cagactgtct 88680
gcttgtagag cccaggagag ctaccagcct tctacccccc cacacctttt tgcaccttcc 88740
ttcaatgcat cccgaacctt cctgggatta tgccaccttc accacaacat cctgaagccc 88800
cagctcagca gccagggtca aaacagttcc aactggaatg aggggctttg tgcgggtgtt 88860
tttaataaac tatcgagaca tttgcataga tttctatgga aacagcatat taggaggctt 88920
aaattagaat ttcaaaaggc tgctatcttg agtgtcgtgg gaagtgggac gagctgtcat 88980
gcttcccagt cttcctgcct gcggtgtgat gacttctttg ttgcctggga tgttgcaaag 89040
aggagtcaaa aaggaggggg aaatctcttt gatttttcct aaaaataatt agactgtgtt 89100
ggcaaatgac catccttcaa acaaaaacag aatccccaga aagcctgact cctaattggt 89160
ttggaggcct cagagacatg agaagggaaa gacagtcctc cgtggtcgag gaaggcgacc 89220
ccttggagac ttttctgcct ggttctcact ctgtggtctg caaacccagc ttcttcttct 89280
ctaacaggtt ttgtcaagag tcaggaagga gagaatgaag agggaagtga aggagagctg 89340
gtggtacggt ttggagagac tcttcccaag gtaatccagg aagaaaatgt gcctggggca 89400
gagggctgca gcagtgcagg ttagtacgca gcactgcagg ctagtgtgca gctacagcag 89460
tgcaggcgag tattcagctg cagcagtgca ggtgagtatt cagctgcagc agtgcaggcg 89520
agtgtgcagc tacagcagtg caggcgagtg tgcagctgca gtagtgcagg ctagtgggca 89580
gcagtgcagg ttggtggaca gctgcagcag tgcaagttgg tgtgcagcag tgcaggttag 89640
tgggcagctg caacagtgca agcgaatatg cagctgcagc agtgcaggtt agtgtgcagc 89700
tgcagctctg taggctagtg tacagctgta gctggtctgc ccagagtaca gctacagcag 89760
gttcatccag actgtagcca tcccagacca aagccataga gaccagctta gacactgtag 89820
tcaagggact tagggtatgc ttttcatcat ggatggatga cctcaatatt cctaagtggt 89880
tttcagccct gctgtgccta ctccccttg gttattacct acttagtgtt cgcagcactt 89940
gagattgaat cgaggctcct gcaagcacta agcaagcacc caaccattga gctatgtccc 90000
tgatttccct gaacactttt tagcagcagg aagttgaggg cagtgctgac atccaggagt 90060
ctgtcactca ccaagccaca tctctagcaa accacagaat atttcaagcc caactggagg 90120
tttgtgacac atttactcag acagggagca ggcctatggt gtcttctcta atgcagccat 90180
agaccttgga ggggtatgag gagagggatg gtactgcaag tctcatttca gacctggctg 90240
ctagaccttc ccttggcctc tgggatatgc aaagaatgga aggaagaatg gaggcaatgg 90300
accagggacc tggaagctca actcaaccca gtttaggaaa tggggtgggg cagggaagaa 90360
aactatctct cacttggctg agcaatgggg tccacctgct ctcttgggca tgaagagcca 90420
agacctgagt ttgcttaaaa gaacatattc agttccgttt tactatgtat ttatttattt 90480
tatgaatgag tgctctatct gcatgtacat ctgcatgcca gaagagggta ttaggttcca 90540
ttacagatgg ttgtgagcca ccatgtggtt gctgggaatt gaatgcagga cctctggaag 90600
agcagccagt gctcttaacc actaagcctc tctccagctc tcctgatgca tgcaaacacc 90660

```
aagactgctg acccccattt acctgtcctg acctgccctc acctgtggac tctaatctgg   90720
tgtctcctct gcagctaaag cccattatct ctgaccccga gtacttactg gaccagcata   90780
tcctgatcag cattaaatcc tctgacagtg acgagtccta tggtaagcat cccaggccag   90840
ggacctcacc gtaccttagc ctcagcctct gaagtgaggg tgatggtggc ttagcctttc   90900
aagaaagcct gtctgccacc acccattgtg ccctgatgct gccacttcct ctgtcccaga   90960
gggacttttc ctagtgggtt accatccctg ctccttgtca cttctgggtg ccacagccag   91020
taacctcctg cttagccttg ctgagtgagc tgcaggtacc atgtaattta atcttagcct   91080
taagaggcag gttctgttcc ccagtgagta aataggtaaa ctgaggctca tagaggtcaa   91140
gtgactcccc ccaggtcaca cagatagtcc tcaactgagc tgtggcttga gcctgagtct   91200
gcctgtgcca ccgtgtctct cgttagtccc tagccagctc ctcagaccca ggtgcagaac   91260
attagcccgt ccctgctgcc aggtcagctt tccagtctgt ctccttttcc tttcagactt   91320
ttccttgtat agagatttta gactgactta cactcttacc tgtttatctg attttgttta   91380
tttcttgtac tgtttcaaac cttgggaagt aagctttccc tatagcaacc agattaacat   91440
ttaataccat ttcttagatg gatttttcta ggatttaatt tttccctcag cattgatctg   91500
tatagagttc tttttcctat gaggcttcag acatttgccc cagattgcta aggcatcgat   91560
tgaaagggtt tgttggcaat ccctgcccgg ccacttgctt caaggagctg tctcccagtg   91620
gtacagccat ctctctgagc atgctctata tccgtctaat gccaagatgt tcctattgat   91680
gctctctctc ctttccccag gtgaaggctg cattgccctt cgcttggaga ccacagaggc   91740
tcagcatcct atctacacgc ctctcaccca ccatggggag atgactggcc acttcagggg   91800
agagattaag ctgcagacct cccagggcaa gatgagggag aagctctatg gtaggtcagc   91860
cagcctcctc ctggctttcc cagaggccac tgcactaagg acatgttctt tctctcagca   91920
aagctatcaa gtatctatga tctgtgcctt cagaatatag cgtggttaga aagtctctag   91980
ggtagcgtgg tgatcctggc actcaggagg ctgaggcgag aagatcgtga gttcaaggct   92040
agtctgggtg acatactgag acgatgtctt taatgagaag aaaacaagtc atctcattta   92100
tctctcacaa ccgccctgtg agtcaagtat ggttatctct acttggcaga tggagaaact   92160
ggagctagat gactggctta atgttgcaaa ctcaggctaa gccaagttca tccagggcca   92220
agcaggagct ggtagacata gtggttgctt gcctgtcatg aaaatagcaa gtgtaggaac   92280
gcacaacatg gggtgcaggg gctccttatg ccagggaggg ctctaagggc cactgcttct   92340
gtgcatgtga ctctcctgtg ccccatgaga gatctttctt ttctctagca ggacagtggg   92400
ttgaacagat agtgggtcat tacctagcaa ccacattcca gaaataaagc ccggggttcc   92460
tttaaactga tttatagggg ttcccataaa cagataagct ctgcatggga agtactctta   92520
tagggtaacc taatagctgt tctaagtttg ttctagtctg acaagaaatg tctgttttat   92580
ccatggctaa gctaacaagg tatttcttga aataacaact gggtcataag tatctgttca   92640
tctccttctg gacccatggc aggcatcatg gcaggcatca tggcaggcaa cagttgccaa   92700
gcgtaaggat tttgaagcct gaaaatcttg cttgatatac ctctcccagt aaccactatt   92760
tactgtccag ctttggcctg caaagagata ccacattgcc atttctaacc catgccgtgg   92820
ctaggatctc acatgttttc tttcctaaat gttctcttcc agactttgtg aagacagagc   92880
gggatgaatc cagtggaatg aaatgcttga agaacctcac cagccatgac cctatgaggc   92940
aatgggagcc ttctggcagg tagacgaagc ttgctaagac ttattacagc tagctgggct   93000
gtgtgaacca gagtccagga gagggtaaag tggagttcag gaaagccaag gtcagaggag   93060
aagaaatgtt gtggcccagg ggacccactc tcctacctca gtttcacatc ctgagttcaa   93120
atcctcatct ctgaaagatc aaagccttgg agaacatttc catgtaggaa gggctcagct   93180
caactatcct ttaatgaaca cctgctatat tcaggaagac acagtgtaaa tacagtatag   93240
tccatgacct acagacttaa caccagccag tggaaggatg tgcagtgcag gaagaaggag   93300
ctatagatga ctgaacaatt ccagggcaaa agatacttct cctgcattca gggaggccag   93360
gaacataagt ggtcaaaagt caccacccat gacttcctag cttcatttct gttgctgtgg   93420
caaatgccct gattaaaagg tagcatgggg agggaaggga tcatttgact tcctcaagcc   93480
cacattacag tctagtgttg tggagaagcc agggcaggaa cctgaagtat cacatcaaca   93540
gtcaagagca cagagaataa acacatgcac ccttgcttgt tctcagttag ttttcttcac   93600
tcttacacag ttcaaggccc agccaaggaa atggtgccac ccacaatgga ctgagtcttc   93660
ctacattaat taacaatgaa gacaagtccc tatagacatg cccacaagcc aaacttatct   93720
agtcaattct tcataaagac tctcttccct ggtgatccta gattgtgtca agagggcatt   93780
tcaaaccaac caacacagag gggtgtagcg gggcatgcct ttacttccag aactcaggac   93840
gtagacgggc tgatcttgtg agttcacggc cagcgttgtc tacacagtga attccaggac   93900
attcagagct acataatgag accctatctc aagaaacaaa ataaaacaaa acacagaaag   93960
actaaccaac acagtgacta acaggaaaga cggattggag agtaacctgg gcattccagt   94020
ctatgttgtt tcctgatatc agaaacaccc ctgtttcctg atgtgctcgt gttagtcctc   94080
```

ctgagatgga gcccattgtg gaacaaggtt tggtgtgtga ctacaaaggt tgcctcacct 94140
ggctctggac tgtctcctcc agttttaaca cagctgccct ccttaccacc tctagcctca 94200
ccctctgtgc cgcacactca ccctgctggt ctgatgcctt ccgtgggatg tgcaaaccaa 94260
catgtcccac gtgactctcc cacagcccca ggcccaccag gttctcattc tgctctcccc 94320
atgtgctgac cttgtcttca gtgaacccca caacagggca gcctctatcc ctgaccacag 94380
ggccttagca tggacctttg cctccacaca gagcatcctc ctctcttttt ctctcttctg 94440
tactccatat aaattcaatc cgcagacagc cttgagtcaa tcttctccag ccatgttcta 94500
cctggtgggt ctccactctc ctacttgtct caacttgatc tcttatttct ctttgccata 94560
ttatacacgt accagaatga gtaaatgaat tgttgtgccc aagaaataaa ggtactgaga 94620
ggctgagaag ctgaggaaca ttttcctcag gctggtgcca gccaagaagg gaagaggctt 94680
caggtgtgcc tccacatctt cctacagggt ccctgcatgt ggtgtctcca gcctcaatga 94740
gatgatcaat ccaaactaca ttggtatggg gcctttgga cagcccctgc atgggaaatc 94800
aaccctgtcc ccagatcagc aactcacagc ttggagttat gaccagctac ccaaagactc 94860
ctccctgggg cctgggaggg gggagggtcc tccaacccct ccctcccaac cacctctgtc 94920
gccaaagaag ttttcatctt ccacagccaa ccgaggtccc tgccccaggg tgcaagaggc 94980
aaggtgagtg tcctctgaat tgtgtgtgtg tgtctgtctg agtgtctgtg tttctgctca 95040
aaagcatccc ttgggggctc ccaagggtga cggcctgaag agggcagagt tgtagtaggt 95100
tctgcccact actttggctt ctgcctgtcc aaaacagttg gtgcaacttg attttaaagg 95160
ggactgggtt gggacttgcc aagaaagcca tcttcttata aaaactgcat ttactcacaa 95220
agacatttct caaaccaatg acaatctatg ccatcctccc cttcctggag ttctcagtgg 95280
gaaagaggtg aggatttctg aatagacacc attatctacc ccaaatctct ctctttcctt 95340
aaaaacaagc attcatgttc tcatatataa ctatctgggg gctggagaga tgactcagtg 95400
gttgagagca ctgactgctc tttcggaggt cctgagttca attcccagca accacatggt 95460
ggctcacaac catctgtagt gagatctgat gccctcttct ggtgtgtctg agaagaatga 95520
cagtgcactc acatacataa gtaagaaaat aactatctgg accctcctga tggaaacata 95580
tattggaaac taggaaccca aatgaaggca gagctgtcat cctacagagg gagccggcca 95640
gaacaggttt aggagcaagg accacacagc ccagagatga agtcctaagc agatatggga 95700
aacattaggt gggaatccca ttcctacagg atgatagatg gccaagtgac atccagacct 95760
agttaacagt gacacagatg tgtctcctcc ccagctgcct tgagcatttt gtggtgacac 95820
cggctctcac attcagcctc ttttctgagt gctgtttctt ttcttccctt ttcagaactt 95880
cagttcctta gcagagctaa ctcatagtaa tcagggaccc gtgctggagc atcacccaca 95940
gtgcggttct cctccagacc ctctaagtca aatgcctata ccaggcctgc ctggggattc 96000
ctgtggagag gcactgacag tcctgtgcct tagctgttag ctgaactagc tgaaaggtgg 96060
gagggcaggt cccttagcca gactgaagtc tacttcctag gagcaaggga gaaatcgcct 96120
tggcatccct ccccggaaat gaggatcaag gtagccatcc agaaggtact gaggtacttg 96180
tttgacaaag gcagtctctt tcgagacccc atgaagcaga gctagagaca gccacaaaga 96240
aagcacaagt tcatggactc agaggttccc agagtggaag tcactgtgtg cttcacacgg 96300
tgaacagagc ttggtggaaa catgtcctct ccagccccag gtgacagcat aggggtagac 96360
agctatcagg gagcgagcta caggactagg tggcaatagc cacccatccc aggttcccca 96420
aggctgcccc aactttagca tttaaaagtc ccatctcctg gaaaacactg cagtcctggt 96480
aagcttggac caccaccatg tgaggtcaag gctgtgtgcc aagaaaggag actctgctca 96540
ggccctagcc ggcatccatg gtctcctgca caagaactga ctctgccccc tggataagag 96600
cttggtttcc tgttgcctat actcaaactt ttttttttt ttgactaaaa ctcgtgagtt 96660
taatctgttt ttcctaacct ccattaggat agcaggtgta ccctgaacac tgctggaagg 96720
atcactttct agcttctatc tcattggctc catcctctgg cctctctatc tctatagctg 96780
agtggatagc cagatgccac acacatgcac aggcacgcac acgcacgcac atgggggggt 96840
ggggagatgg gggggagtt gcttgcttct gtctaccaca atgccctcag aggccagaag 96900
aggacattgg attcccaaga actgcagtta catacagttg tgagttacta tgtagatgcc 96960
aggagtcgaa cctggtaact ctggaagaag agcaaatact cttaatcaca gaaccaactc 97020
ttcagcccgg catggtgttg catgccttta atcccagcac tcagcagaca gtgaggcggg 97080
cggatctgtc agttctatat caaccaggga tacgtagtga gccctggtgt ttaaaaagaa 97140
aaattctcaa ggcagaatcc atgtaaaaat tattccctgg aaaataagta attcggaggc 97200
atttctgtgg tgcatgttat atgatcacta ctgtagttag ttccagagca ttcttatcac 97260
ccctaaatga aacccagagg catgaagcaa ctactcccca tgtcctgact tctgaacata 97320
tttcctcccc atctcctctc cttttcccca tggttccaga cctggggatc tgggaaaggt 97380
ggaagctctg ctccaggagg acctgctgct gacgaagccc gagatgtttg agaacccact 97440
gtatggatcc gtgagttcct tccctaagct ggtgcccagg aaagagcagg agtctcccaa 97500

```
gatgctgcgg aaggagcccc cgccctgtcc agacccagga atctcatcac ccagcatcgt   97560
gctccccaaa gcccaagagg tggagagtgt caaggggaca agcaaacagg cccctgtgcc   97620
tgtccttggc cccacacccc ggatccgctc ctttacctgt tcttcttctg ctgagggcag   97680
aatgaccagt ggggacaaga gccaagggaa gcccaaggcc tcagccagtt cccaagcccc   97740
agtgccagtc aagaggcctg tcaagccttc caggtcagaa atgagccagc agacaacacc   97800
catcccagct ccacggccac ccctgccagt caagagtcct gctgtcctgc agctgcaaca   97860
ttccaaaggc agagactacc gtgacaacac agaactcccc caccatggca agcaccgcca   97920
agaggagggg ctgcttggca ggactgccat gcaggtatgt tgagctgtat gtatatgggt   97980
gtatatgcat atgtgtgcac gcatgcatct gtgtgtgtgt gtatgtgtat gtgtatcctg   98040
ggatgtgctc tgtgaagaag ggctaaacct ggggtttgta cctggcatct gttcctgttc   98100
ctgactggtc ctacagggca tctgcagtga ttcttgagag cgacagagag gagctgttga   98160
cacctgtaga gtaagggagt ctaagtcatt tatattttag tcttatcatc cttaattagg   98220
gtatgcactt aaaacattcc tggtgtttta gagacctcag gcagaatatc cccttcctat   98280
agaatagatt tattgtagag caaaactaac aatgacttaa atacacacac acacagagtt   98340
gagactggac atgatagtac accgctataa tcccagcaca ctgtacatct agagcaagag   98400
gattggaagt tcaaggcaag cctgggctaa atggtgagac agacaacagc agcagtgaga   98460
gagctaaggt tatatagctc agtgggagag tgcacaggaa gcctccagaa agcccttagg   98520
tgaatctcca gtatcacaag gaacagagtc tagaactaga ctgctggcca ttccattagc   98580
atacttattg cagcattgag ggtcagactt ggggccttac actaggcaag cactgtacca   98640
cagtcctgtg ttggatttaa atcccaagtc tccactgcac ggctctatgc ctttttaaac   98700
tgtcaaaaag aaactactag tgccagccat tgtgtagatt tatggggaga aaaagattat   98760
gccatatact gttattgttt gtgatcattt ttacagttat cactcatttt gaaactatat   98820
acaacatata tctatctcat aggtactgtg tttgtaatat atgcatgaat taacatataa   98880
ggcacttagg acagtgtggc tctttcacat taattagcct gacctaactt acagcactag   98940
gggtcaaatc tggcaaggcc ctgatagccc tttctgttct catcttacag tgagctgctg   99000
gtgatcggag cctggaggaa cagcacaaag cagacctgcg cctctctcag gatgcctctc   99060
tcaggatgcc tcttggagga cctcctgcta gctcttcttg cctagcttca agtcccaggc   99120
tgtgtatttt ttttcaggaa acggcctcac ttctctgtgg tccaagaagt gtgctgctgg   99180
ctgccacact gtgcggcaga tgctaaagct ggatgacaaa cgcacgccat acagacagca   99240
gacagcggca ctgggtctca gaacttggat tcctgggcct tcttccagtc gccgttttaa   99300
agaaaggaac taacggagct gctcatccga gggtgaagat ataaataata atattattaa   99360
taataataac agtcaggtgc catgtgctgt gttaagtgct ttatgaacat ttgtcgggct   99420
ggcctccagt gctgaggtgc cagtcagcct gaaccctatg cccaggccca ctaatcccaa   99480
atggtgggtc ctgagatgtt tttaaaaagc attaaagaaa accatcggtc tcttagagct   99540
aaccggccgg gctctactgc agggacccga acagtctgca tggctaagtg gcacaaggag   99600
cctggccctg tccagcttca gagatccaag ctgctttttg ctggggttct gtcacaggcc   99660
tgatcctctt ggtttttatg gggtttcaag tctgccagag tcagaaatca gctctaactc   99720
gccagtgaag agatctggcc ttaacttaag ccagccacgt caggcccctg ctgagcctat   99780
ggaccaataa atactccccg tgccactgga ggtgggcagc tatcaccata ccctgagttg   99840
ggccaagccc accccacccc taccctgcaa catttctgat gtactgagga agagtctcca   99900
ccatagtccc caagggctga gttctccagc ctgctatcag ggaaggtgag cattggtccc   99960
aggctctcaa aatagtgcag cctcttcttc ccaagctctg gggtgcaccc tgtgtccttg  100020
gttaccagga gactagggtt gtgatatctt ttcttgtctt gctttttgat atatcaggat  100080
taatgtagga aaccagacct agattattca ggagagtagg tatatccccc gtgtttccca  100140
```

<210> 2
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 2
tactcctcag caagagtagc tgg 23

<210> 3
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
Primer

<400> 3
gctgaacttg tggccgttta cgt 23

<210> 4
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
Primer

<400> 4
cttctatagc cttcccaagc c 21

<210> 5
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
Primer

<400> 5
ctcgtaggtc tcacaggaag 20

<210> 6
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
Primer

<400> 6
cctgctggat tacattaaag cactg 25

<210> 7
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
Primer

<400> 7
gtcaagggca tatccaacaa caaac 25

<210> 8
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
Primer

<400> 8
ggcgttctct ttggaaaggt gttc 24

<210> 9
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
Primer

<400> 9
ctcgaaccac atccttctct 20

<210> 10
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
Primer

<400> 10
ttgctgcacg agggctcaga atc 23

<210> 11
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
Primer

<400> 11
tccgattctc atgctctggc ttg 23

<210> 12
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 12
cagccctgtc tttgccacgt ttg 23

<210> 13
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 13
tccactggat tcatcccgct ctg 23

<210> 14
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 14
cttcctcttg caacagagaa ccc 23

<210> 15
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 15
actcaacgtc cactttgaga tgc 23